PART V

Between Firms

21

Vertical Integration and Market Structure

Timothy Bresnahan and Jonathan Levin

1. Introduction

Vertical integration occupies a central role in organizational economics. Williamson (2005) calls it the "paradigm" problem for explaining the distribution of firms and markets in modern economies. In this chapter, we review research on vertical integration decisions and their consequences, and offer some perspective on the current state of knowledge.[1]

Our discussion bridges two very different literatures. Research in organizational economics generally treats vertical integration as an efficient response to contracting frictions. This approach is often associated with Coase (1937) and Williamson (1971). Research in industrial organization has taken a complementary approach, emphasizing patterns of integration at the market or industry level. Some of this work, following Stigler (1951), draws attention to scale and scope economies as rationales for integration, while other strands emphasize strategic motives and the idea that integration can be a tool for consolidating or extending market power.

Theoretical work in the first tradition argues that certain features of transactions create particular problems for arm's-length contracting. These can include difficulty anticipating future contingencies, ambiguity in the nature of tasks and decisions to be carried out, the need to use specific assets, or an inability to measure and verify transaction outcomes. In Section 2, we describe some of the "building-block" models that link these types of transactional frictions to problems of hold-up, decisionmaking externalities, and incentive distortions, and offer explanations for when and why internal organization might lead to more efficient outcomes.

We thank the volume editors, Robert Gibbons and John Roberts, for helpful discussions and unusual patience, and the Ewing Marian Kauffman Foundation (Bresnahan) and the National Science Foundation (Levin) for their support.

1. Readers may also be interested in the surveys by Gibbons (2005), Joskow (2005), Klein (2005), Shelanski and Klein (1995), and Lafontaine and Slade (2007). Perry (1989) is an earlier survey of industrial organization research on vertical integration. Helpman (2006) and Antras and Rossi-Hansberg (2009) review recent work on multinationals and cross-country integration decisions.

The ideas in these models have been the starting point for a long sequence of empirical studies dating back at least to Monteverde and Teece (1982), Masten (1984), Joskow (1985, 1987), and others. The typical strategy in the empirical literature has been to relate observed integration decisions to measures of contractual frictions, or more often, proxies for these frictions. In a very few cases, an attempt is made to link the integration decision to economic outcomes (e.g., productivity, investment, and innovation), or observed transaction costs (e.g., the incidence of disputes).

Many studies along these lines report statistically significant correlations between integration decisions and proxies for contracting difficulties. In Section 3, we discuss the extent to which these findings can be viewed as strong supportive evidence for the theory. We highlight a number of difficulties in relating the evidence to the theory. One is that the relevant theoretical variables (e.g., the importance of noncontractible decisions or the marginal returns to specific investments) can be inherently hard to measure (Whinston 2003). In other cases, the sources of variation that identify the parameters of interest appear to reflect important omitted variables. Finally, even if one accepts the directional results, the lack of evidence linking integration decisions to realized transaction costs makes it hard to assess the quantitative importance of the effects.

Empirical research also has faced a tension between testing specific predictions derived from the theory and trying to explain broad industry patterns of integration. Most contractual theories of integration aim to isolate particular mechanisms. The models tend to focus on simple environments, for example, two parties structuring a supply relationship with the surrounding market held fixed. Thus, the clearest empirical analogues are marginal decisions at the firm boundary, such as whether a firm should make or buy a particular input. Moving from this level of analysis to explain the organization of an entire industry requires additional steps, because the availability of suppliers or the existence of functioning markets or contracts needs to be treated as an outcome of the analysis rather than a fixed parameter.

This higher level industry perspective on vertical integration has been adopted in much of the industrial organization literature. Here, decisions about integration are often considered in the context of broader models of industry structure, competition, or technological change, but typically with much less attention to contractual issues and the ways in which vertical integration might differ from sophisticated contractual arrangements. In Section 4, we discuss this research, which has an empirical component as well, consisting most often of industry case studies or historical analyses of industry evolution. Using the case studies to communicate the key analytical themes, we distill three reasons industry structure and firm structure are interdependent.

The first relates to the presence of scale or scope economies. Many theories of industry structure are framed in terms of trading off the successful exploitation of scale economies versus the benefits of competition. If that trade-off calls for a large number of distinct firms in one or both of the industries where firms might be vertically integrated, the contracting problems emphasized in some organizational economics theories decline in importance. This leads to models in which market size or thickness is an important determinant of vertical integration.

Second, competitive strategy can be a key factor in vertical integration decisions. A rich set of models in industrial organization emphasize the use of vertical integration as a way to raise entry barriers in one or both of the associated markets. Another class of models focuses on mixed

structures, where a firm in industry A might be vertically integrated with a firm in industry B while also serving other firms in B as a disintegrated market supplier. Such structures create complex strategic problems of supplying, or being supplied by, one's competitor. These models lack foundational theories regarding why integration, rather than complex contracts, is needed to implement a particular strategy, but they nevertheless capture trade-offs that commonly arise in industry case studies and are discussed frequently in competition policy.

Third, the industrial organization literature has taken up the special case of vertical integration with complementary innovators. This literature deals with the trade-off between the benefits of coordination associated with integrated production and the benefits of getting multiple diverse draws, which favors separation into multiple independent inventors. The distinction has organizational implications, because contracting to coordinate ex ante calls for identifying and selecting a partner ex ante. Market organization of innovation can use ex post market selection. We use the production of scientific research as an example to illustrate this trade-off and how it brings together organizational and industry-level ideas.

We conclude our review by noting that the two literatures on vertical integration have remained largely separate despite the potential benefits of cross-pollination. We provide some suggestions for how insights from organizational economics might inform traditional analyses of market structure, and how a broader industry-level perspective at times might enrich organizational economics.

2. Contractual Theories of Vertical Integration

In his famous essay on the nature of the firm, Coase (1937) proposed to study why and how economic activity divides between firms and markets. He argued that firms exist to reduce the costs of transacting through markets. Their size and scope depends on the extent to which they offer a transaction cost advantage. Therefore, a decision to vertically integrate rather than sourcing inputs or selling outputs should reflect the respective transaction costs of internal and market organization.

To go further, one needs an explanation for why market transactions are inefficient and why the inefficiencies cannot be remedied with better contracts or pricing mechanisms. One also needs to explain why integration might yield an improvement in some cases but not all. Finally, a satisfactory theory should match observed patterns of integration and offer some hope for empirical testing and measurement. Essentially all contractual theories, from transaction cost economics to formal models of hold-up, decision rights, and incentives, reflect these goals.

2.1. Transaction Cost Theories

The transaction cost theory of integration was developed by Williamson (1971, 1975, 1985), Goldberg (1976), Klein et al. (1978), Joskow (1985), and others beginning in the 1970s. The theory expands on Coase's original idea by describing a wide set of transaction inefficiencies and potential organizational responses.[2]

2. Tadelis and Williamson (this volume) provide a more detailed account of transaction cost economics. See also Joskow (2005) and Williamson (2010) for recent overviews.

A starting observation is that market contracts are inherently incomplete. Parties can plan for some contingencies but not every one of them. Thus, in a great many transactions there will be room for opportunistic and inefficient behavior as the transaction proceeds. The concern may be especially severe when complexity or uncertainty make it difficult to specify contractual safeguards or when parties cannot walk away without incurring substantial costs.

Transaction cost theory argues that integration can be an effective response when these features are present. One reason given by Williamson (1971) relates to decisionmaking. When a dispute arises in an organization, it can be settled by a senior manager. In contrast, a dispute between separate entities must be resolved by negotiation or litigation. To the extent that managerial authority makes it easier to resolve disputes or make adaptations, integration can be the efficient response to uncertainty or contractual incompleteness (Williamson 1971, 1979).

A related rationale for integration is that it might mitigate potential hold-ups. If parties anticipate the possibility of future haggling or disputes, they may have little incentive to make specific investments for fear the investment could be wasted or expropriated (Klein et al. 1978; Williamson 1985). To the extent that integration allows a firm to protect specific investments, it again can be an efficient response to contractual incompleteness.

These stories provide a motive for vertical integration but not a corresponding cost. Williamson (1975, 1985) suggests a related but distinct set of inefficiencies inside organizations. These include low-powered incentives, and rent-seeking and informational bottlenecks that arise in managerial hierarchies. An implicit assumption in transaction cost theory is that these problems are relatively insensitive to the complexity, uncertainty, or specificity of particular transactions. So integration becomes optimal when hazards are more severe, but nonintegrated production may be preferred for relatively straightforward transactions.[3]

An important point in many transaction cost analyses is that the set of potential organizational responses to transaction frictions is very large. In addition to full vertical integration and spot contracting, responses can include contracts of varying duration and specificity, joint ventures, partial ownership through financial investment, partial control through seats on a board or voting arrangements, or shared ownership of assets or intellectual property. In fact, part of the appeal of transaction cost economics is that it offers such a broad and inclusive framework for matching organizational responses to transaction frictions. From an empirical perspective, however, this richness poses some challenges.

The sharpest predictions of the theory are that such transaction characteristics as specificity, uncertainty, and complexity should favor vertically integrated production. Perhaps the canonical example of specificity is a coal-fired electricity plant located next to a coal mine (Joskow 1985), which is also a good example to illustrate the subtleties of empirically testing transaction cost theories of integration. Coal-fired electricity plants are expensive to build, have a long lifetime, and are designed to burn particular types of coal. Before investing, the owners want to ensure a steady and reliable input supply. There are, however, a variety of ways to do this. Some plants vertically integrate, while others sign long-term contracts with the neighboring mine. There is relatively little evidence that the integrated plants outperform their counterparts

3. Williamson (1985) also emphasizes that the frequency of a transaction can be important, in the sense that one-time transactions are less likely to be integrated than repeated transactions.

(Kerkvliet 1991). So even in cases where specificity is apparent, vertical integration may be only one of several possible responses and not necessarily the most effective.

This same point can be seen in another classic example from the transaction cost literature: the acquisition of the Fisher Body Corporation by General Motors (Klein et al. 1978; Coase 1988, 2000; Klein 1988, 2000, 2007; Langlois and Robertson 1989; Casadesus-Masanell and Spulber 2000; Freeland 2000). Fisher Body supplied metal car bodies to GM in the early days of the auto industry. In 1919, GM acquired a stake in Fisher and the companies signed a contract committing Fisher to supply its output exclusively to GM. The contract was for 10 years, on cost-plus terms. After it was signed, production expanded as the demand for cars grew, and in 1926, GM purchased the remainder of Fisher Body, dissolving the contract and integrating the firms.

The question of why GM decided to integrate Fisher Body has been the subject of extensive debate. Klein et al. (1978) interpret the episode as a canonical example of a hold-up leading to integration. They observe that the relationship involved substantial specific investments in manufacturing and human capital. They also argue that rising demand raised Fisher's opportunity costs, so that they began to hold up GM by, among other things, using a relatively costly manufacturing process. This left GM with no option but integration. Coase (2000) and Freeland (2000) provide a very different interpretation. They point out that Fisher and GM had a tight relationship. A Fisher brother sat on GM's board, and GM paid for the construction of at least one Fisher plant. Rather than a hold-up arising from the cost-plus contract, Freeland and Coase argue that the cause of integration may have been a dispute about whether to expand an existing plant or build a new one that fit GM's expansion plans.

The episode demonstrates the subtleties in trying to formulate sharp empirical tests based on transaction cost theory. The historical research makes it clear that even before integration, Fisher and GM were closely intertwined. It is not obvious that things would have proceeded much differently if the companies had integrated in 1919 or not at all. Moreover, although different accounts are at odds over many issues, they are all consistent with the broad claims of transaction cost economics. In this sense, the same inclusiveness of the theory that makes it useful for organizing historical accounts also makes it difficult to reject.

A somewhat different issue, but one that turns out to be important for applied work, is that transaction cost theory tends to treat individual transactions in isolation, holding fixed the characteristics of the transaction and the surrounding market conditions. The difficulty in applied work is that transaction characteristics and market structure are often best viewed as endogenous variables. For instance, firms may have a choice about whether to use standard or customized inputs or to cooperate with other firms in their industry to create standardized contracts or technologies. We return to this point in more detail in Sections 2.5 and 4.3.

2.2. The Property Rights Model

The property rights model (Grossman and Hart 1986; Hart and Moore 1990; Hart 1995) is one of the first formal theories of integration that makes precise assumptions about the limits of contracting.[4] The model focuses on how integration changes the incentives to make specific

4. Segal and Whinston (this volume) provide an overview of the property rights model as well as many extensions and related models.

investments. In the property rights theory, a firm is defined as a collection of physical assets. Decisions to integrate upstream or downstream are associated with shifts in asset ownership. An example would be a manufacturer that purchases a supplier and takes control of its physical plant or one that integrates downstream and takes ownership of distribution outlets. Contracts are assumed to be incomplete. When open issues arise, ownership strengthens a party's bargaining position. As a result, it has a stronger incentive to undertake ex ante specific investments.

The standard version of the property rights model has three stages. The relevant parties first decide who should own which assets. Other decisions, or some of them, are left to be decided later. Then the parties make investment decisions. The investments are specific—some of their value is lost if the relationship breaks down. After investments are sunk, the parties bargain over the outstanding terms and trade. Bargaining is assumed to be efficient, so the surplus-maximizing trade is made regardless of asset ownership.

Although bargaining leads to an efficient trading outcome irrespective of the distribution of assets, ownership does affect the terms of trade. Because it conveys the right to use an asset in the event of breakdown, ownership of an asset means that a party gets a larger share of the bargaining surplus. Under the typical assumptions about payoffs, this leads to sharp predictions about investment incentives. A move toward integration increases the investment incentives of the acquiring party but lowers the incentives of the acquired party. Determining whether integration is optimal, and which party should be the owner, requires a comparison of the costs and benefits, which can be complex.

Although the hold-up mechanism seems similar to transaction cost theory, the property rights model leads to distinct empirical predictions (Whinston 2003). For instance, transaction cost theory predicts that integration will be a response to specificity, where specificity refers (somewhat loosely) to a gap between the value of the ongoing relationship and the value of the parties' outside alternatives. In the property rights model, the gap per se is not consequential for predictions about ownership (Holmström and Roberts 1998). Instead, the relevant quantity is the degree to which ownership affects the marginal returns to noncontractible specific investments.

Perhaps because of these subtleties, the specific predictions of the theory have not received a great deal of empirical attention. Whinston (2003) emphasizes the difficulty in measuring or proxying for the marginal returns to investment. Moreover, ownership in the model conveys *residual* rights of control—that is, the right to make decisions that are not otherwise specified in a contract. In many settings, it is not obvious how to characterize or measure the importance of these rights. Similarly, the model focuses on investments that are both noncontractible and inalienable. Identifying important investments that are inherently nontransferable (e.g., investments in human capital) and yet important enough to drive vertical integration decisions can be tricky.

Holmström (1999) also has emphasized another issue with the property rights model, which may help explain why it has been more influential as a theoretical building block than as a framework for empirical work. Taken literally, the model suggests that individuals (e.g., the ones investing in human capital) should own the productive assets. Clearly, this is not the right frame of reference for thinking about large organizations that own many assets and coordinate many activities, where shares of ownership are often widely dispersed and separation of ownership and

control is common.[5] One could associate the actors in the property rights theory with larger entities and still think about hold-up of specific investments as being a rationale for vertical integration, but in the context of the model, one then needs an explanation for why investments cannot be shared or contractually specified.

2.3. Decision Rights Models

One major contribution of the property rights model is that it offers a precisely formulated argument for why and when ownership conveys the right to make decisions. Many earlier accounts (e.g., Simon 1951) simply assume that managers enjoy a degree of authority to make decisions.[6] In the property rights model, a firm's owner has to bargain with employees, who are free to walk away. If they do, however, they leave the firm's assets behind. This can give a firm's owner some degree of bargaining power over the firm's employees. In contrast, a supplier that walks away takes productive assets with it. Of course, in the property rights model, the key decision rights are never exercised because of the efficient bargaining assumption, so ownership matters only through ex ante investment incentives.

An alternative, and in some ways simpler, modeling approach assumes that bargaining is not efficient, or that there is no bargaining at all. Examples of this approach include Hart and Holmström (2010) and Baker et al. (2011). In these models, ownership is associated with the right to make certain residual decisions (the presence of physical assets is no longer very important). Roughly, these models predict that ownership should be allocated in a way that ensures more efficient ex post decisions (see Segal and Whinston, this volume).

In Hart and Holmström (2010), for instance, there are two business units that can benefit from coordinating their actions. Integration means that the units are placed under a single manager. The manager can force the units to coordinate, but (by assumption) may fail to respect the private preferences of the divisional employees. If instead the units operate under separate ownership, the respective managers will cater to their employees, but the units may fail to coordinate. There is a trade-off between the coordination benefits of internal organization and the incentive or appropriation benefits of nonintegration.[7]

This type of decision rights approach hearkens back to a much older, although less contractually grounded, set of models in industrial organization. For instance, in the classic double

5. The separation of ownership and control in large organizations relates closely to many of the issues raised in this chapter, though we will not pursue it here. (Berle and Means (1932) is an early and classic reference.)

6. This view has been controversial. For instance, Alchian and Demsetz (1972: 777) write:

> It is common to see the firm characterized by the power to settle issues by fiat, by authority, or by disciplinary action superior to that available in the conventional market. This is delusion. The firm does not own all its inputs . . . [An employer] can fire or sue, just as I can fire my grocer by stopping purchases from him or sue him for delivering faulty products.

7. Similar trade-offs also appear in the literature on internal capital markets (e.g., Stein 1997). A related model in Hart (2009) explores how integration might provide a way to avoid costly hold-ups, as in the transaction cost account of Klein (1996). In Hart's model, vertically related parties write a contract to trade at a future date. But they are unable to get the price right for every contingency. A situation can therefore arise where one party might threaten to walk away, triggering a (costly) renegotiation. The optimal ownership structure therefore is one that minimizes the likelihood of a hold-up.

marginalization model of Spengler (1950), a nonintegrated upstream supplier will choose its markup without regard to the profits of a downstream producer. The resulting trade is inefficient. Integration helps resolve the inefficiency, because pricing decisions are coordinated in a way that leads to a jointly profit-maximizing transfer price.[8]

The double marginalization model and related models of incentive conflicts in vertical supply chains share the feature that integration shifts authority from separate managers to a single coordinating entity. What it leaves unexplained is why integration does a better job of restoring efficiency than a sophisticated nonintegrated contract would (Tirole 1988). Nevertheless, it is a workhorse for applied economists, because of its clear predictions (e.g., the potential inefficiency will be larger and the incentive to integrate greater when there is substantial market power both upstream and downstream). In this sense, the approach taken in recent decision rights models may be a promising framework for empirical research, particularly for studies where it is possible to observe decision processes and outcomes in integrated and nonintegrated situations.

2.4. Incentive Theory Models

A complementary set of vertical integration models focus on how integration decisions shift the financial incentives of workers and managers. As a simple example, we can consider a firm contracting with a sales agent. Suppose for the moment that nonintegration is associated with the agent having high-powered incentives—for instance, a contract under which the agent buys the product at a fixed price and keeps the proceeds from its sales. Conversely, integration means that the agent becomes a salaried employee with low-powered incentives. Whether integration is optimal will depend on how important it is to provide incentives for the sales agent and whether monitoring can substitute for high-powered financial incentives.

This type of agency model has been the dominant paradigm in research on franchising arrangements and related distribution integration decisions. It has several attractive features. First, internal organization often does seem to feature low-powered incentives relative to market-based exchange. Second, the basic logic of incentive theory models is familiar, and the relevant quantities—the sensitivity to incentives, the difficulty of measuring or monitoring behavior, the degree of risk aversion, and so forth—are in principle objects that can be proxied or quantified.

One element that is missing, however, is why integration should be equated with low-powered incentives in the first place. Holmström and Milgrom (1994) provide an answer. Their argument is that financial incentives generally are part of a broader incentive environment. Other elements might include decisions about asset ownership, or the set of tasks that are permitted or excluded. Holmström and Milgrom show that under certain conditions, asset ownership by the agent (e.g., nonintegration) can be complementary to providing high-powered financial incentives. Conversely, taking assets away (integration) makes it desirable to use low-powered incentives. From this perspective, integration is likely to be associated with broader shifts in incentives and other organizational variables, even if asset ownership per se is not the primary driver of incentives (see also Holmström 1999).

8. Related models focus, for instance, on the inability of nonintegrated firms to coordinate investment or marketing decisions, or to price complementary products efficiently. Tirole (1988) and Perry (1989) provide extensive coverage of these types of models.

The incentive system model is particularly useful for thinking through shifts in incentives and organizational design that might accompany integration decisions. For example, privately funded start-up firms in Silicon Valley often compensate their workers and managers with equity or options. The work atmosphere can be intense. These firms attract workers and managers who are drawn to the high-pressure, high-powered incentives of starting a new business. If a large firm acquires the start-up, it may seek to align the start-up's activities and efforts with its own. Measuring alignment can be difficult, so the logic of the model suggests that an acquisition will be followed by a shift to lower-powered incentives and perhaps the subsequent departure of some of the original managers and employees.

The challenge that arises in making the theory operational for empirical work is that its predictions are again very rich. Many outcome variables should move together as a response to underlying shifts in the environment. As a result, most of the research linking agency considerations to integration decisions has tended to focus on simpler, although perhaps less interesting, predictions.

2.5. Capabilities Theories

The theories we have described focus mainly on decisions at the boundary of the firm and often deal with cases where firms integrate to internalize the transfer of some tangible good or service. In contrast, research on business strategy and organization theory frequently emphasizes the idea that firms may seek to expand or acquire other firms to leverage their internal capabilities or exploit superior management capabilities (e.g., Penrose 1959; Cyert and March 1963; Wernerfelt 1984). These theories rest on two premises. The first is that organizations develop certain capabilities or know-how that is embodied by managers and employees or in organizational routines. The second is that capabilities or knowledge cannot easily be traded or shared across firm boundaries.

These ideas have received less attention in economic theory, although there is a range of work that speaks to related issues. In the Lucas (1978) model of firm size, managerial talent is a scarce resource that can be leveraged by creating hierarchical organizations. Lucas uses the model to derive predictions about the size distribution of firms in the economy. Subsequent work has adapted his model to study the optimal organization of hierarchies (e.g., Garicano 2000). Clark (1985) and Bloom and van Reenen (2007) provide evidence that managerial practices can be an important factor in explaining productivity differences across firms.[9]

These theories address the first piece of the capabilities view—that firms differ in managerial know-how, and this difference is important for production—but not the second. That is, they do not address why integration, as opposed to contractual arrangements, is required to leverage managerial talent. One natural hypothesis is that managerial authority tends to coincide with firm boundaries, an idea that is captured in the model of Hart and Holmström (2010) discussed in Section 2.3, and is a common presumption in less formal analyses.

9. These papers look at comparable firms and have an observable definition of "better" managerial capabilities. As Clark (1985) points out, to understand complementarities between different organizations, one needs a definition of managerial capabilities that includes the task to be carried out as well as effectiveness at pursuing that task.

A complementary line of research argues that vertical integration allows information flows that cannot occur across firm boundaries (Malmgren 1961). For instance, Arrow (1975) considers a model in which information can be transmitted within an integrated firm but not between disintegrated firms. Again, the model sidesteps the question of why transmitting information or knowledge across firm boundaries might be difficult. One potential justification is that employees who interact regularly in a firm develop a body of shared knowledge that facilitates further communication.

Another difficulty with sharing knowledge or information across firm boundaries is that once information is shared, it cannot be rescinded, creating an appropriability problem. Indeed, Cohen and Levinthal (1990) argue that the ability to absorb information can itself be an important organizational capability. In Section 4.4 we argue that identifying situations in which firms have an advantage in coordinating the production and dissemination of knowledge is particularly useful for thinking about the organization of innovative industries.

3. Empirical Evidence on Vertical Integration Decisions

Since the 1980s, many empirical studies have attempted to test or apply contractual theories of integration. The studies vary in the details but share some features. The general approach is to predict observed integration decisions using transaction characteristics identified in the theory. The integration decisions lie at the firm boundary—whether to buy certain inputs, to use a contracted sales force, to franchise retail outlets, or instead to bring these services in house. These studies focus on integration decisions at the firm level, holding fixed the nature of the goods and services being contracted over, the contracting environment, and the surrounding market structure.

These papers also share common measurement challenges. Testing contractual theories of integration requires empirical measures of concepts, such as the degree of specificity, the importance of noncontractible investment, and the potential for opportunism and hold-up. Testing also requires variation in their incidence. The variation might be across firms or across the transactions of a single firm. But its source needs to be plausibly exogenous to the integration decision itself. It is also necessary to identify the set of possible contracting outcomes. This undertaking is complicated by the possibility, noted above, of a wide range of contractual solutions to a given problem. Most studies try to capture moves closer to or further from integration with some sort of binary classification and avoid fine distinctions among contractual arrangements.

Virtually all studies we discuss report statistically significant correlations between integration decisions and proxies for theoretically relevant transaction characteristics. This correlation has been viewed as quite favorable for contractual theories. A more skeptical perspective, however, is that these analyses often do not provide a very sharp test of the theory. In some cases, the empirical proxies used to capture complexity, specificity, or the potential for opportunism do not have an obvious relationship to the theoretical concepts. In other cases, the source of identifying variation seems to be earlier or concurrent decisions made at the firm or industry level.

Perhaps the main limitation of this work, however, is that very few studies provide any quantitative (and sometimes not even qualitative) sense of the transaction costs or incentive

distortions associated with different contractual forms. As a result, it is frequently unclear whether the difference between integration and nonintegration is really that important in any economic sense. In Section 3.6, we argue that this last concern requires combining evidence on integration decisions with specific ex post measures of transaction costs or performance, and that this approach is a key direction for future research.

3.1. Sourcing Complex and Specific Inputs

Some of the earliest empirical studies look at industrial make-or-buy decisions in search of evidence that firms might seek to integrate the production of complex or specific inputs. Two classic examples are Monteverde and Teece (1982) and Masten (1984).[10] Both studies relate make-or-buy decisions to empirical proxies for hold-up and opportunism and thus are usefully representative of a much larger empirical literature.

Monteverde and Teece (1982) studied how Ford and General Motors source auto components. They used a probit regression to relate sourcing to measures of transaction specificity and complexity. A component is "specific" if the procuring firm is the exclusive purchaser of that component. An engineering cost estimate is used to proxy for complexity, or technical know-how. Because many components consist of parts procured in different ways, Monteverde and Teece defined a component as "made" rather than "bought" if the cost fraction of the component produced internally exceeds a given threshold. Their main finding was that both specificity and complexity are positively correlated with internal sourcing.

Masten (1984) examined a single large aerospace project that required nearly 2,000 inputs. Masten also constructed measures of specificity and complexity, in his case on an input-by-input basis. Here too, an input is specific if the general contractor is the exclusive user. Inputs are classified as complex or not complex, based on a survey of the procurement team. Masten found that both measures are highly correlated with sourcing. In fact, more than 90% of the specific and complex inputs were sourced internally versus less than 1% of the nonspecific and noncomplex inputs.

These pioneering studies highlight some of the challenges in testing contractual theories of integration. One challenge is to find empirical measures of transaction frictions that correspond closely to the conceptual ideas in the theory. For instance, both studies use measures of technical or engineering complexity as proxies to indicate that purchasing a component is complicated. The relationship is clearly imperfect. Computers are very technically complex, and yet most firms readily source computers from external vendors. From a theoretical perspective, it would be better to define complexity in terms of communication or information about a transaction. Are there many possible contingencies? Does performance involve a great many details that are difficult to communicate or incorporate in a contract? Is it difficult to identify the responsible party if problems arise? Is there substantial room for asymmetric information? These notions each require a different approach to make them operational. For example, the idea that complex transactions may involve asymmetric information calls for an empirical treatment recognizing that one party may have superior information at the time of contracting. This idea, so important

10. Related work on industrial or government procurement decisions includes Walker and Weber (1984), Masten et al. (1989, 1991), Crocker and Reynolds (1993), and Levin and Tadelis (2010).

in the rest of the economics literature on asymmetric information, appears essentially nowhere in the empirical literature on vertical integration.[11]

Defining "specificity" is also problematic. Monteverde and Teece (1982) and Masten (1984) use what is most naturally viewed as an outcome, the demand for an input by only one customer, to proxy for a primitive, namely the specificity of the investments associated with the exchange relationship. The fact that the input has a sole user may reflect some fundamental and fixed attribute of the input (and in a study that focuses narrowly on a single firm, that may be the obvious source of variation on which to focus), but it may also reflect broader market forces. For example, variation in the efficient scale of production relative to demand needs may play a large role in determining whether an input is classified as specific. If the input is efficiently produced at large scale relative to the demand of individual users, we might expect to see several firms using the same input, even if it is produced by only a single supplier, creating the type of bilateral market power emphasized in the theory. If production of the input is flexible, users may adopt customized versions, irrespective of whether the market to supply the input is highly competitive or very concentrated.

Note that the important part of this argument for analytical purposes is not the specific point about scale. It is the movement of the argument from firm to market. In the firm-level theory, an input is made rather than bought because producing it requires specialized equipment, knowledge, or a specialized location, or because the input itself has no alternative use, and moreover because this specialization leaves parties open to opportunism if the input is bought instead of made. In the market-level alternative theory, some inputs are sold to multiple buyers because that is the efficient market organization, and this causes their use to be less specific. This arrangement might be optimal because of scale economies in production, because the market is organized in a way that facilitates the pricing and distribution of standardized products, or because regulatory or legal uncertainty favors standardized transactions.[12]

These problems illustrate the difficulty of distinguishing aspects of a transaction that are primitive from those that follow from design decisions or market equilibrium. Baldwin and Clark (2005) have tried to tackle this problem from an engineering perspective. They argue that within a production process it is possible to identify the engineering interfaces that are more or less amenable to a market transaction. One suspects this approach could prove powerful in certain instances, particularly if it is possible to measure performance along a chain of production (e.g., failure rates or delays).

11. Many of the most successful approaches to asymmetric information are based on datasets in which the econometrician, by gathering data later, can observe an outcome that only one party could observe at the time of contract (e.g., Einav et al. 2010). This approach is beginning to be deployed in the few studies of vertical integration that examine outcomes of this sort. Below, we look at a paper by Arora et al. (2009) that studies pharmaceutical development for licensed (i.e., contracted) versus in-house chemical entities. They interpret the observed lack of difference in success rates between licensed and in-house developed drugs as evidence against any "lemons" theory of asymmetric information in the licensing market.

12. One can also raise issues of causality with reference to a single firm. If the market supplies a small set of input varieties, a producer may be able to tailor its design to use these varieties rather than a specific input. Similarly, Novak and Stern (2009) point out that if there are complementarities in production, manufacturers may internally source clusters of inputs if they anticipate difficulties in coordinating design changes that require large-scale modifications.

We have emphasized the difficulties in measuring transaction frictions, but difficulties in measurement also arise on the left-hand side of transaction cost regressions. In the case of auto manufacturing, U.S. manufacturers traditionally relied on a partial outsourcing model in which suppliers compete for contracts each year. Japanese firms adopted a very different approach involving close relationships with a set of closely coordinated suppliers (Asanuma 1989). The economic and engineering problems faced by firms in this industry in different countries are not all that different, but history and broader institutional issues led these firms to adopt different organizational modes. Depending on the study country, one might arrive at very different views of what it means to buy rather than make a component.

3.2. Specificity, Market Thickness, and Hold-up

Empirical researchers arguably have had more success in linking integration decisions, or at least the use of more complete vertical contracts, to measures of market structure (e.g., the availability of multiple suppliers in the upstream market). Two emblematic studies are Stuckey (1983) and Joskow (1987). These studies get around many of the problems discussed above by identifying settings where parties make large investments in specific assets—in each case, specialized production facilities—and there appears to be substantial potential for hold-up in subsequent price negotiations stemming from the limited availability of trading partners.[13]

Stuckey (1983) focuses on aluminum manufacturing. The primary input in this process, bauxite, is very costly to transport. Thus, it is efficient to locate aluminum refineries close to bauxite mines, and indeed the largest refineries are mine-mouth plants. Because refineries and mines require substantial sunk investments and have long lifetimes, Stuckey argues that integration is a natural response to the threat of costly breakdowns. He documents that integration of co-located mines and refineries is indeed common. Hennart (1988) provides an interesting comparison: the production of tin. Relative to aluminum, tin production has much lower capital and transport costs, less specialization of refineries, and also much less vertical integration.

Joskow (1987) studies coal-fired electric utilities. He points out that utilities located in the eastern United States have relatively easy access to multiple coal suppliers, while utilities located in the West have fewer options. As a result, western utilities are more exposed to costly hold-ups and are more likely to take contractual precautions. Geographic variation in market structure therefore provides a source of identifying variation. The dependent variable in the study is the duration of supply contracts, as vertical integration is relatively uncommon. Joskow shows that utilities in the East have relatively short contracts, whereas those in the West can last 15 years or more. Over such a long duration, the value of energy can be expected to vary widely. Thus, the costs of long-term contracting might be considerable, especially if the contractual price only poorly tracks the spot prices that would have prevailed (Joskow 1988). Presumably this cost is weighed against the benefit of protecting investments in choosing contract duration.[14]

13. Acemoglu et al. (2010) take a less direct approach to assessing the relationship between investments and integration, but one that is motivated by property rights theory, studying the relationship between technological intensity and integration in British manufacturing.

14. A related line of work uses variation across firms within an industry, or across local markets, to test whether integration decisions are related to the number of available suppliers or to upstream concentration. See, for instance, Lieberman (1991), Ohanian (1994), or the additional studies discussed in Section 4.

In the aluminum and coal examples, asset specificity is closely related to market structure. In Stuckey (1983), co-location and high transport costs ensure that refineries have few alternative suppliers. In Joskow, firms in the West cannot rely on a thick spot market and competition to avoid hold-up. In each case, a concentrated market structure (i.e., few ex post options) provides an empirical proxy for potential opportunism.

The availability of alternative suppliers can depend a great deal on the relevant time horizon. A firm may have many options if it plans well in advance, but few on short notice, creating what Masten et al. (1991) refer to as "temporal specificity." Pirrong (1993) argues that differences in the incidence of temporal specificity can explain organizational arrangements in ocean shipping for different commodities. Grain and fertilizer are at one extreme. They can be carried using general-purpose vessels and have liquid spot markets. The thickness of the market leaves little room for haggling over prices, and both commodities are commonly shipped under spot contracts. In contrast, iron ore requires specialized vessels to ship, and the number of mines and steelmakers is limited. Integration and long-term contracts are the norm. Either steelmakers own their own shipping subsidiaries, or they secure a shipper for a long period of time, with the agreement typically negotiated prior to the construction of the specialized ship.

3.3. Decision Rights and Coordination

One of the central ideas in transaction cost theory and in models of the optimal allocation of decision rights is that integration might help firms adapt to contingencies that are anticipated in a broad sense but not in specific detail. Testing this proposition calls for a setting in which such contingencies arise frequently and with enough similarity to facilitate statistical analysis, but with sufficient variation to prevent easy contracting in advance. Several studies have made progress by looking at integration decisions in transportation industries, where scheduling poses a coordination problem that often has to be solved in real time in a fashion that allows adaptation to changing circumstances and events.[15]

Forbes and Lederman (2009) study the relationships between major U.S. airlines and their regional affiliates. Some carriers own their regional affiliates, while others use long-term contracts. The organizational structure is revised infrequently, but numerous issues (e.g., scheduling) must be resolved on an ongoing basis. Forbes and Lederman focus on the possibility that weather disruptions might be a particular source of conflict. They argue that when there is a disruption, a carrier may want to adapt local flights to keep their network running as smoothly as possible. Regional carriers may not internalize broader problems on the network. Forbes and Lederman use cross-route variation in weather patterns to proxy for the incidence of conflict. They find that major carriers are more likely to use wholly owned regional operators on routes with worse weather. Integration is also more common on routes that are more intricately linked to the network, where the need for coordination is arguably larger.[16]

15. Gil (2007) provides a related analysis of adaption in the movie industry, where the market for any given movie is initially uncertain but realized over time, leading to a question of how long theatres should continue to exhibit the movie when the box office receipts are being shared with (disintegrated) movie distributors.

16. Forbes and Lederman (2009) do not provide much direct evidence on the costs of integration that encourage carriers to sometimes rely on contracts with nonintegrated affiliates. They hypothesize that arm's-length

In a subsequent paper, Forbes and Lederman (2012a) tighten this story by relating integration decisions to ex post measures of performance, in this case flight delays. They find that integrated regionals experience fewer delays than nonintegrated ones, particularly on airport days with adverse weather conditions. The finding is consistent with the idea that the national airline, not the regional carrier, disproportionately bears the costs of delay and that opportunistic behavior by the regional carrier hinders adaptation in certain contingencies. Although delays do not translate immediately into dollar costs of disintegrated supply, the paper is unusual, and admirable, in completing the link from plausibly exogenous market conditions to integration decisions to ex post performance.[17]

Problems of adaptation also feature prominently in studies by Lafontaine and Masten (2002), Nickerson and Silverman (2002), and Baker and Hubbard (2003, 2004) on organizational arrangements in the trucking industry. In trucking, costs are reduced by minimizing driving distance and maximizing the likelihood of full loads. These features mean that efficient matching of trucks and loads can create substantial value, and often the matching needs to be done in real-time, based on the location of trucks and loads. One conjecture is that when load matching is more important, we might expect a greater prevalence of trucking employees driving company-owned trucks rather than independent operators contracting for loads.

Consistent with this broad hypothesis, Nickerson and Silverman (2002) find that employee drivers are more common in "less-then-truckload" (LTL) carriage, where they argue there is more need for coordinated scheduling. Baker and Hubbard (2003, 2004) find a similar relationship between the use of employee drivers and difficulty in finding backhauls, where again the potential for miscoordination and haggling may be greater. In an interesting extension, Baker and Hubbard (2003) look at the adoption of information technology in the form of on-board computers. They point out that on-board computers had ambiguous implications for organization form. On the one hand, by allowing companies to track driver locations, they facilitated load matching and more efficient capacity utilization, favoring an integrated structure. On the other hand, they allowed for improved monitoring of driver activities, facilitating arm's-length contracting.

The Baker and Hubbard (2003, 2004) papers and Pirrong's (1993) work on shipping raise the interesting question of whether vertically integrated control structures have a natural advantage in solving complex coordination or matching problems. In transportation, scheduling and load-matching problems are frequently tackled using optimization models that incorporate various constraints on allocation and dispatch. Economists immediately understand that solutions to these types of problems have a market analogue, for instance with multipliers on different constraints corresponding to implied prices for different routes. What is the incentive to shift from market control to operations research control? Pirrong points to the problem of having sufficiently thick markets to generate reliable prices. Another rationale might be the

contracting keeps the employees of the regionalfirm at a distance from the generally quite powerful unions of the parent carrier.

17. In a further related study, Forbes and Lederman (2012b) analyze an industry shift from revenue sharing contracts with regional carriers to capacity utilization (or "per flight") payments. They show that this shift was associated with the adoption of flexible regional jets and argue that the change in contract was needed to reduce haggling with the regional carriers over what flights to fly.

difficulty of solving complex scheduling and matching problems in a decentralized fashion, although arguably this type of problem might be addressed with more sophisticated market mechanisms.

3.4. Agency, Incentives, and Monitoring

Empirical researchers have also explored the extent to which incentive contracting theories of integration have predictive power. The large literatures on retail franchising decisions and sales force integration are typical in this regard.[18] Here, researchers have attempted to link organizational decisions by the parent firm—such as whether franchise outlets are to be company owned—to the importance of franchisor and franchisee effort and to difficulties in monitoring this effort. Because direct measures of the marginal returns to effort and the ability to monitor typically are hard to come by, most researchers rely on proxies. For example, studies have found that more remote outlets (e.g., those in rural areas, without neighboring outlets, or farther from headquarters) are more likely to be franchised as opposed to company owned (e.g., Brickley and Dark 1987; Minkler 1990; Kehoe 1996). Other studies report correlations between franchise ownership and such variables as employee-to-sales ratio or the experience required of new franchisees.

A common issue in interpreting these findings is that (once again!) the empirical proxies may not relate very tightly to a particular theory. For example, the cost of monitoring remote outlets may be higher, making it desirable to provide higher powered revenue incentives. Alternatively, remote outlets also may be less integral to the chain's reputation, making it less important for the parent to exercise control. The set of individuals interested in obtaining a remote franchise and the potential customers at remote outlets also may be different from nearby franchises.[19] This ambiguity can lead to multiple explanations for a reported correlation.

Some studies have been able to mitigate the proxy problem by identifying cases where there is compelling cross-franchise or cross-location variation in the nature of tasks. Shepard's (1993) study of gasoline stations in Massachusetts is a good example. Shepard views the choice of organizational form for gasoline stations as resolving a trade-off between resolving double-marginalization problems in pricing and providing incentives for the station operators. She argues that incentives are likely to be more important at stations that offer ancillary services, such as auto repair or full-service gasoline. Consistent with this logic, Shepard finds that the stations offering ancillary services are less likely to be company owned.[20]

18. On franchising decisions, see, for example, Caves and Murphy (1976), Brickley and Dark (1987), Norton (1988), Minkler (1990), Lafontaine (1992), Lafontaine and Shaw (2005), and Yeap (2006). Studies of sales force integration include Anderson and Schmittlein (1984), John and Weitz (1988), and Regan (1999).

19. Other studies proxy for franchisee effort using such measures as the employee-sales ratio (Norton 1988) or the amount of franchisee experience required (Lafontaine 1992). Related work adopts proxies for the importance of franchisor effort, such as advertising expenditures (Lafontaine and Shaw 2005) or the number of years a chain went before franchising any outlets (e.g., Lafontaine 1992). It is not entirely obvious why the variation in these variables should be unrelated to other factors influencing franchisee ownership.

20. Yeap (2006) takes a similar approach to studying franchising decisions by fast-food chains. Her central finding is that restaurants where food preparation is done on site and table service is supplied are more likely to be franchised. See also Slade (1996).

Studies of retail organization also face difficult problems of selection. For instance, a common hypothesis in research on franchising is that parent companies should tend to own outlets where revenue is highly variable, as a way to shelter franchisees from risk. The evidence, however, is highly inconclusive (Lafontaine and Slade 2007). Of course, finding a convincing proxy for risk is not easy, but even apart from this measurement problem, one may be concerned about selection. For instance, franchise operators may vary substantially in their risk attitudes, and those with greater risk tolerance (or perhaps greater wealth) may be drawn to the more risky outlets and also prefer to have ownership.[21]

A broader issue with the contractual literature on franchising is that it does not provide much evidence on whether outlet ownership has first-order operational effects. After all, a statistical relationship between incentives and ownership structure does not necessarily establish that the incentive problems under one organizational form are large. From an incentive perspective, a company-owned McDonald's might be radically different from a franchisee-owned McDonald's, or it might be pretty much the same. To assess the difference, one needs to know whether there really is substantial variation in effort or behavior across franchises and whether it covaries with organizational form. Absent such information, it is difficult to infer whether agency considerations are really the central issue driving franchising decisions or whether such issues as capital structure and financing are more important. The best way to address these questions would be to incorporate direct measures of outlet performance into the analysis.

3.5. Knowledge and Information Transfers

In contrast to empirical work on transaction specificity and incentives, the literature that tries to link vertical integration decisions to knowledge and information flows is less well developed. One likely reason is that the relevant variables—organizational capabilities or know-how, managerial ability, and internal processes—can be difficult to measure, a prerequisite for empirical work. (Bloom and van Reenen's (2007) survey measures of management practices are an interesting and promising exception.) As discussed earlier, the theory is also less sharply organized in terms of generating specific testable hypotheses.

One area where there has been some progress is the study of organizational structures in the pharmaceutical industry. Knowledge creation in pharmaceuticals happens in the early stages of drug discovery and in subsequent stages of development and clinical trials, and there are further stages of production and distribution. Several studies have found evidence for economies of scope at various stages of the chain, which they attribute to firms developing specialized capabilities.

Henderson and Cockburn (1996) focus on the first stage of drug discovery, using data on development projects gathered from large pharmaceutical firms. Their dataset is unusual in that it contains information on projects that do not succeed as well as those that do. The number of patents provides a measure of success. By this measure, projects taking place in firms with larger overall research programs appear to be more successful, and Henderson and

21. The same concern can be voiced about Shepard's (1993) study to the extent that the operators of auto repair stations are relatively entrepreneurial and may prefer to be owners—although Shepard does try to provide auxiliary evidence that weighs against selection bias.

Cockburn argue that this can be attributed in part to scope economies. Their work suggests that complementarities across research projects may help explain the boundaries of pharmaceutical firms.

Subsequent to their study, the industry underwent a major shift toward early-stage drug discovery taking place outside the large pharmaceutical firms, in academic laboratories or smaller entrepreneurial firms. This shift creates a potential market or contracting friction. Relatively few firms appear able or willing to push drugs through to commercial use, whether because of scale economies or differences in capabilities at the later stages of drug development. As a result, a licensing market has emerged, where drugs invented in one firm are transferred for development or commercialization in another. A natural concern is that the licensing market might suffer from problems of expropriation or asymmetric information.

Arora et al. (2001) investigate whether licensing transactions might be subject to a "lemons" problem of asymmetric information. They gather a dataset that includes ex post outcomes that might not have been forecastable by (all) parties at the time of contract. They look at the later stages of drug trials and commercialization for licensing agreements signed before phase I clinical trials. It appears that, contrary to the lemons hypothesis, success rates do not differ between products that are out-licensed and those retained in-house for further development. Arora et al. also conclude, perhaps less surprisingly, that there are economies of scale and scope at the firm level in the later stages of drug development.

Concerns about organizational capabilities are also the key to Azoulay's (2004) case study of outsourcing in the pharmaceutical industry, where firms increasingly rely on contract research organizations to perform clinical trials for new drugs. Azoulay looks at how the decision to outsource varies with the nature of the trial. He argues that the incentive systems and capabilities of large pharmaceutical firms and contract research organizations are fundamentally different. Contract research organizations are organized to efficiently gather and analyze large quantities of routine data, whereas pharmaceutical firms are better equipped to manage trials that require the production of new knowledge rather than just new data. The data appears to be consistent with this story. For instance, Azoulay finds that studies that involve more academics, and hence may be more skewed toward knowledge production rather than data production, are much less likely to be outsourced.

These interesting results suggest connections between knowledge transfer and organizational capabilities and integration in one specific industry. Even so, one might argue that these papers are not yet fully testing what is arguably the most interesting hypothesis related to integration in innovative industries. Specifically, smaller, newer firms might have the advantage of flexibility and diversity (particularly with respect to new drug ideas), whereas larger established firms might have superior commercialization and regulatory skills. These differences might be behind the pattern of licenses, alliances, and entry and growth of firms of different types.

3.6. When and How Do Vertical Integration Decisions Matter?

We have already highlighted one of the main limitations of the empirical literature. By and large, it focuses on the decision of whether to vertically integrate, but not on the implications of this decision for economic outcomes or realized transaction costs. This focus leads to questions

about when integration has first-order incentive effects (e.g., in the case of franchising), effects on decisionmaking or adaptation (e.g., in the case of sourcing different types of inputs), or important effects on innovation and productive efficiency.

Quantifying the effects of integration decisions poses some challenges. One difficulty, of course, is the econometric problem of selection. If firms choose to integrate when they think it will lead to productivity benefits, finding the right variation to identify the effect of integration on productive efficiency is going to be difficult. In the case of coal contracts, for example, one might like to measure the costs and benefits of longer contracts holding fixed the market environment. Most of the utilities with short-duration contracts, however, are in the eastern United States, whereas those with longer duration contracts are co-located or are in the western United States.

A second difficulty with measuring the effects of integration is that most empirical studies focus on borderline cases, where integration is a close call. Of course, integration may have both a large efficiency benefit and a large efficiency cost, but this is not guaranteed. Thus, a more promising strategy may be to look at decisions that are complementary to integration. For instance, as emphasized by Holmström and Milgrom (1994), small variations in the environment might lead firms to make large changes to a cluster of policies if there are important complementarities.

These difficulties notwithstanding, there is a variety of evidence that sheds some light on differences across integrated and nonintegrated firms. One example comes from research on manufacturing productivity.[22] For instance, Hortaçsu and Syverson (2007) compare integrated and nonintegrated concrete and cement plants. They find that integrated plants are larger and more productive in the cross-section, that plants that become integrated are more productive, and that the event of becoming integrated is associated with an increase in productivity. They argue that the source of these productivity advantages may be tied to better logistics coordination, a benefit that is associated more with size than with vertical integration per se.[23]

Further evidence along these lines is provided by Atalay et al. (2012), who use establishment-level data from a broad range of industries. They again show that integration, plant size, capital intensity, and labor productivity are positively correlated. The paper also includes a striking finding about commonly owned plants in vertically related industries. There is relatively little transfer of physical goods from one plant to another. Instead, plants within the same firm in vertically related industries are overwhelmingly involved in merchant operations, in which they transact across the firm's boundary.

22. Less work has been done on behavioral or operating differences between integrated and nonintegrated manufacturers. For instance, Mullainathan and Scharfstein (2001) study the capacity decisions of vinyl chloride monomer manufacturers, some of whom are integrated downstream into the production of waterproof plastic, of which this monomer is an input. They present evidence that nonintegrated vinyl chloride monomer producers react more strongly to market demand for the monomer, whereas integrated producers focus on internal demand needs, and they argue that this might be explained by differences in managerial focus and attention.

23. Related work by Van Biesebroeck (2006) on automobile manufacturing finds that component integration is associated with higher productivity, especially in plants that have adopted flexible manufacturing techniques. His results also are consistent with the view that integration facilitates greater coordination in the production process.

This observation strongly suggests that much common ownership of vertically related plants is driven by the sharing of intangible assets (e.g., managerial skills, know-how, or other information) or is related to the financing of the firm.[24] This general story is also consistent with work on large diversified firms, such as that of Schoar (2002), who provides evidence that plants in diversified firms are more productive than those in single-segment firms.[25] This sort of evidence pushes us away from a primarily operational or commodity-flow understanding of vertical integration.

Relating these papers to the theory calls for telling where exactly the trade-offs arise in making integration decisions. To what extent can managerial know-how be leveraged, and where do the benefits drop off? Does an increase in scale or scope feed back into productivity? These analyses also do not make much contact with the types of contractual concerns on which theorists have focused. They also differ from those discussed above in that there is little explicit measure of what happens at the firm boundary or at the divisional boundary within a vertically integrated firm. Instead, an inference is drawn from firm scale and scope. Thus, the interpretation in terms of the underlying theory is necessarily indirect, although a natural trend is toward theories that involve spreading a fixed managerial input over a range of production activities (e.g., Lucas 1978).

The productivity and incentive consequences of vertical integration have also received attention in research on healthcare organizations. In the United States, healthcare systems have become significantly more integrated over the past 25 years, through partnerships between general and specialist physicians, physicians and hospitals, and hospitals and specialty care units, among others. Although one explanation for this integration is that physicians and hospitals are attempting to improve their bargaining power vis-à-vis insurers and exercise market power, another line of thought holds that integrated systems are able to successfully resolve contractual inefficiencies and can provide more efficient, coordinated care.

The current literature seems to be broadly inconclusive (Gaynor 2006), but some studies offer interesting specific findings, some of which are consistent with contractual theories. Afendulis and Kessler (2007), for instance, look at the problem of integrating diagnostic services and treatment. They find that in the case of coronary artery disease, patients seen by cardiologists whose practices also provide surgical treatments were more likely to receive angioplasties; these patients received generally more expensive care but did not have better health outcomes. Afendulis and Kessler interpret this result as consistent with moral hazard or incentive contracting theories of integration.

In more recent work, Afendulis and Kessler (2011) look at how changes in Medicare reimbursement policy affected skilled nursery facilities that were and were not integrated with

24. In an effort to make this finding more direct, Hortaçsu and Syverson (2007) also examine the white collar (nonproduction) workers in the plants they study. Their inclusion might tie the inference closer to a theory about management. However, many of the white collar workers in manufacturing plants are not, typically, managers of the manufacturing production process, but rather managers of the market or nonmarket processes by which intermediate inputs flow into the plant and outputs flow out of it.

25. More surprisingly, however, she finds that existing plants become less productive following an increase in firm diversification. She hypothesizes that this result also can be explained by management, to the extent that a new acquisition leads to a shift in managerial focus away from existing lines of business. This interpretation is consistent with another systematic merger fact, which is that many mergers are followed by spin-offs of the acquired business (Ravenscraft and Scherer 1987).

hospitals. They find that a shift to prospective payment (essentially fixed-price reimbursement rather than cost-based reimbursement) led to significantly greater declines in healthcare costs for the integrated facilities. The mechanism that enables integrated hospitals to better respond to changes in external financial incentives is not entirely clear, and indeed, studies have reached mixed conclusions on whether hospitals that are more broadly integrated generally have lower costs (Ciliberto and Dranove 2006; Cuellar and Gertler 2006).

Ciliberto (2006) uses integrated and nonintegrated hospitals to offer a more specific test of hold-up theories, by looking at the decision to invest by offering outpatient and diagnostic services. These services rely on referrals from primary care physicians, so the hypothesis is that hospitals that are integrated with their physicians might be better positioned to internalize referrals, which might support investments in service provision. Ciliberto indeed finds that vertically integrated hospitals added more of these services following integration. The effect is strongest in areas where HMO penetration is higher, which is roughly consistent with the idea that the returns to integration might increase with the prospect of hold-up.

What are the efficiency consequences of this finding? The perspective Ciliberto (2006) takes is that the potential for hold-up leaves nonintegrated hospitals investing too little in outpatient and diagnostic services, and integration resolves the inefficiency. An alternative interpretation, closer to the view in Afendulis and Kessler (2007), is that offering these services provides a mechanism for physicians to capture rents. These rents could arise from failures of the physicians to act as the perfect agents of patients in making referrals or from problems in the setting of prices paid for these services by third-party payers.

The example illustrates the difficulty of moving from even cleanly measured behavioral effects to efficiency or welfare interpretation in a complex market and organizational environment, such as health care. Nevertheless, we expect that this topic will, and should, receive a great deal of attention in coming years.

4. Vertical Integration and Market Structure

A recurring theme in our discussion of the empirical literature on vertical integration is the connection between integration decisions and market structure. This link has not been the focus of contractual theories of vertical integration, but it is a major focus of research in industrial organization. In this section, we distill some of the key ideas and relate them to work in organizational economics.

The theories we examine bring several different sets of ideas to the foreground. One set of ideas relates to the determination of horizontal market structure, typically based in firm-level costs or in strategic interaction among firms. In some circumstances, cost efficiencies or strategy decisions that impact horizontal structure also affect vertical organization. Another set of ideas concerns the creation of institutions supporting market exchange and the importance of these institutions for supporting disintegrated forms of production. The final set of ideas pertains specifically to innovative industries and the efficient organizational and market structure for coordinating and spurring innovative activity.

Apart from their focus on industry-level outcomes, the industrial organization theories we describe differ from the organizational theories in paying far less attention to the foundations of contract theory. They tend to assume, without spelling out exact reasons, that integration

offers more control or a greater ability to coordinate decisionmaking than a market contract does. For this reason, one can think of the two sets of theories as complementary inputs in a full explanation of vertical integration, even if this complementarity has not always been fully recognized.

4.1. Scale Economies and the Extent of the Market

An important line of research in industrial organization relates vertical integration to fixed costs and market size. This approach has its roots in Adam Smith's dictum that the division of labor is limited by the extent of the market (Young 1928; Stigler 1951; Bresnahan and Gambardella 1998). The central idea is that vertical integration decisions are closely connected to the horizontal structure of the constituent industries.

This point can be understood in a simple model that involves the production of two complementary inputs subject to scale economies. Suppose the market for the final product is so small (relative to the scale economies) that it can support only a single firm in each of the two input markets. Because of the horizontal structure in each of the two markets, the vertical relationship between the two producers is bilateral monopoly. If this relationship leads to inefficiencies because of haggling, hold-up, incentive distortions, or any of the other reasons explored above, the efficient vertical relationship calls for integration. The same problem, however, may not arise if the product market expands. Suppose that final demand becomes sufficiently large (relative to scale economies) to support a competitive market structure in both layers. With this horizontal market structure, the bilateral monopoly problem need not arise, and the efficient vertical structure may involve disintegrated specialists.

Note that this argument blends two very different bodies of theory. It requires a theory of horizontal industry structure, drawn from industrial organization, and a theory of bilateral contracting, drawn from organizational economics. Like most theories we discuss in this section, the elements are simple. But they focus attention on the joint determination of firm and industry structure and the interaction between ideas from the two fields.

Stigler (1951) developed this basic idea into a theory of industry evolution. He argued that in the early and innovative phases of an industry, firms have to be vertically integrated, because there are no markets for the relevant inputs—the costs of organizing those markets, he assumes, are higher than the costs of coordinating the production of the inputs within the firm. As the industry becomes larger, what had formerly been internal inputs are supplied by new, vertically disintegrated industries. In addition to the trade-off between efficient horizontal scale and vertical market power, Stigler's theory adds the idea that market institutions to support disintegrated trade are themselves endogenous and have to be developed over time.

Many historical accounts of particular industries resonate with Stigler's theory. One example is Rosenberg's (1994, 1998) account of how design procedures for chemical processing plants evolved over time. Initially, manufacturing firms that operated chemical processing plants undertook plant design themselves. Plant design is complex and represents a plant-level fixed cost, but much of the knowledge needed to design a single plant can be used to design other plants. Thus, the potential existed for a separate disintegrated design industry that could allow the acquisition of this knowledge to be shared across multiple plants.

As Rosenberg (1994, 1998) points out, this potential was not immediately realized, because it required the creation of a new market boundary—identifying exactly what services specialized engineering firms would provide in plant design. Ultimately, however, specialized engineering firms did emerge. Their emergence subsequently changed the degree of scale economies in chemical manufacturing, because firms no longer needed to have many plants to spread the fixed costs of maintaining a plant design capability. The change in scale economies helped facilitate a less concentrated industry structure in a number of distinct chemical processing industries (Arora et al. 2001).

Another example in which one can see evidence for Stigler's (1951) ideas comes from the integrated circuit industry. These circuits must be designed and then manufactured. A few decades ago, the creation of an interface between computer-aided design and computer-aided manufacturing permitted the design and manufacture to be done either by the same firm or by different firms. As the manufacturing facilities are very large, and some products have very short production runs, the creation of the interface led to a great deal of vertical disintegration. Firms that specialize in design but do not manufacture are now able to contract with firms that specialize in manufacturing but not in design.

Scale economies play a role here, because many products with small markets may be produced in the same large manufacturing facility. There are also individual products with very large markets, such as Intel microprocessors. Firms selling the products with the largest markets, such as Intel, are vertically integrated. This permits them to strategically align manufacturing process improvements with product design improvements, presumably with less contractual frictions than might arise in a situation of bilateral monopoly between layers of production. So the variation in the extent of the market across products (rather than over time) generates patterns of integration consistent with the Stigler theory.

In addition to historical case studies of individual industries, some attempts have been made to statistically test relationships between vertical integration and varying measures of scale economies or horizontal concentration. Holmes (1999) takes an interesting approach by asking whether vertical disintegration is more prevalent in geographic areas where manufacturing industries are localized. For each plant, he measures localization as the amount of employment at neighboring plants in the same industry. Roughly, the idea is that in areas where an industry is localized, there is the potential for specialized suppliers to thrive and support a disintegrated vertical production structure. Holmes finds that vertical disintegration (measured by the share of purchased inputs in production) is indeed higher in areas where there is extensive local employment in a given industry, broadly consistent with Stigler's hypothesis.

An alternative cross-sectional approach has been to test whether industries that are more highly concentrated or have larger firms are also more vertically integrated. The Levy (1984) paper is an older but carefully executed study along these lines. Levy looks at the relationship, across three-digit industries between vertical integration (as measured by value added divided by sales) and covariates, such as average firm size, demand growth, and concentration. He reports that industries with larger firms, more demand growth, and more concentration are all more likely to feature vertical integration (see also Elberfeld 2002). Even with careful execution, however, these types of correlations are very difficult to interpret as sharp evidence for or

against any particular theory, because there are so many factors that can influence cross-industry relationships between vertical integration and other industry characteristics.[26]

One of the elaborations of Stigler's (1951) general approach to thinking about vertical integration occurs when complementary goods are produced with very different levels of scale economies. This can lead to the common industry structure in which one or a few firms supply a general specialty or platform input that is complementary to a potentially large number of products or inputs produced by other firms (Stigler 1951; Bresnahan and Gambardella 1998). A typical trade-off here is that any efficiency benefits that might come from the platform provider integrating into complementary markets must be weighed against the potential for reduced competition.

For instance, consider a setting in which differentiated products rely on the same complementary platform. Standard models of market structure in differentiated products industries give rise to an equilibrium in which firms enjoy some market power to cover the fixed costs of product differentiation, and a trade-off occurs between achieving scale economies in a particular product versus having product variety and competition across products. If the horizontal industry equilibrium has many more firms than scale economies in the complementary layer permit, vertical integration that forecloses the platform to competing products may have the effect of eliminating the benefits of competition and differentiation in the market where they are feasible.

An example considered by Stigler (1951) is the production and transportation of manufactured goods. Efficient industry structure in production is likely to involve many distinct firms, whereas parts of the transportation system are subject to increasing returns and thus are supplied by only a few firms. Both competition and product diversity drive the production layer toward multiple firms. Although some manufactured goods (e.g., iPads) may be produced by only a few firms and others (e.g., lag bolts) by many, in the aggregate we expect a large number of firms producing a wide variety of goods. In contrast, scale economies play a much larger role in determining the equilibrium industry structure for transportation by limiting the number of railroads, canals, or roads that can link any two places as well as the number of ports or airports that can be built at any given city.

Note that there is some tension between the industrial organization theories of market structure in the types of industries we are considering (i.e., those with differentiated products) and the organizational economics theories. The industrial organization theories take certain firm boundaries as given. For instance, they offer no explanation for why a single firm could not control and produce a wide range of differentiated products. Instead, they begin with the assumption that no firm could integrate all manufacturing processes, even if it remains an open question whether each manufacturing firm owns ships or railroad cars. Although this argument can be seen as merely one of common sense, it implicitly relies on some notion of organizational diseconomies of scale or scope.

26. Acemoglu et al. (2009) take a somewhat different approach by asking whether firms in countries that are more financially developed are more or less likely to be vertically integrated. They find no direct relationship between financial development and the prevalence of integrated firms, but they do find a higher prevalence of integration in countries that have both greater financial development and higher contracting costs.

This gap has been met partially by organizational economics theories that explain organizational diseconomies of scale. Williamson's (1985) suggestion that bureaucratic decisionmaking is one disadvantage of large firm, Simon's (1947) idea that larger firms must undertake more communication, and Radner's (1992) idea that more complex and larger teams will make worse decisions have all received some exploration in this regard.[27] These questions, which have received only modest attention in organizational economics, become particularly important when we consider the joint determination of firm and industry structure.[28]

4.2. Role of Strategy in Determining Industry Structure

The simple framework discussed above suggests a relationship between horizontal concentration and vertical integration that is driven by efficiency considerations—the size of scale economies relative to the size of the market. What it neglects is the possibility that firms might make strategic efforts to limit competition. An industry can be very concentrated and feature considerable market power because of scale economies or because business strategy limits the number of active firms (e.g., by affecting the potential for entry).[29] Indeed, that organizational economics theories have focused so much on efficiency rationales for vertical integration suggests that we might want to reformulate Coase's crack about monopoly theories: nowadays if an economist finds something—a business practice of one sort or another—that she does not understand, she looks for an efficient contracts explanation![30]

Once we allow for the possibility of strategic investments, the relationship described above between market size and concentration can become less clear. Consider the class of industries that feature what Sutton (2007) has called "endogenous sunk costs." These are industries in which firms can undertake strategic investments (e.g., in technological innovations or brand building) to make their products more attractive. If the investment opportunities in endogenous sunk costs are sufficient (in a sense made precise in the literature), then industry structure remains concentrated, even as demand grows without limit. The link from increased market size to reduced horizontal concentration to vertical disintegration is broken.

A very large literature on industrial organization also has pointed to the potentially important role of vertical integration or vertical contracts themselves in helping to limit competition or more generally influence horizontal market structure. A classic example might be a monopoly supplier of an input that integrates or signs an exclusive contract with the supplier of a complementary input to maintain monopoly power in the first market or create it in the second. The literature has debated two questions about this type of situation. First, is it even possible for

27. See, among others, Simon (1951), Calvo and Wellisz (1978), Williamson (1985), Radner (1992), Anton and Yao (1995), and McAfee and McMillan (1995). Some evidence on the source of scope diseconomies in technology industries can be found in Bresnahan et al. (2012).

28. Another partial solution comes from the industrial organization theories themselves. Models of horizontal disintegration through competitive entry call for separate firms in order to have competitive outcomes. Vertically integrating all firms in a competitive industry with the same complementor would defeat the competitive purpose.

29. See, for example, Berry and Reiss (2007) and Sutton (2007).

AU: Please supply the page number for the Coase (1972) quote in footnote 30.

30. 30. Coase's (1972: 00) original line was: "If an economist finds something—a business practice of one sort or another—that he does not understand, he looks for a monopoly explanation."

the monopoly supplier to raise prices by integrating? Second, is it anticompetitive, and as such, should antitrust policy seek to prevent this type of strategic foreclosure?[31]

Whinston (2007) provides a lucid and thorough treatment of these issues, which go well beyond the scope of this chapter. We note, however, that this literature—although potentially highly germane if one wants to understand the joint determination of firm and industry structure—has by and large not connected with organizational economics rationales for integration. Indeed, it often treats vertical integration and restrictive vertical contracts in roughly a parallel fashion. It is true that in formulating antitrust policy, the potential organizational efficiency benefits of allowing a contract or merger to go through are sometimes weighed against the competitive costs, but generally these are treated as distinct phenomena.

We also note that in addition to strategic theories in which firms integrate as a way to gain market power in anticompetitive fashion, there are arguments for why strategic integration might be welfare enhancing. For instance, Teece (1986) has suggested that firms can sometime integrate to consolidate "good" market power. If a firm has an invention in an area with poor patent protection, for example, it might vertically integrate into a downstream business with entry barriers and avoid losing the return on its innovation through competition.[32] Another class of theories, discussed for instance in Perry (1989), consider how vertical integration can facilitate certain forms of price discrimination that might be efficient (e.g., by allowing a supplier to integrate downstream and sell its input at different prices to the integrated downstream producer than it does to the general market).

4.3. Market Institutions and Vertical Integration

The point made by Stigler (1951) that market institutions need to be invented before a wide variety of firms can participate in them remains only partially explored in organizational economics. Distinct literatures have taken up this question, many of which raise interesting issues about vertical integration and market structure.

A textbook benefit of a well-functioning market is that the price mechanism coordinates the needs of many parties and determines an efficient use of resources. If matching is less efficient, a buyer and seller might agree to go around the market. This idea is captured in Carlton's (1979) model of supply assurance as a rationale for vertical integration. He considers a setting in which market frictions upset the functioning of the spot market for inputs, so that buyers cannot be assured of reliable supply. If the problem is severe, buyers want to integrate upstream, or

31. The Hart and Tirole (1990) paper is a classic reference on the first question, which turns on the question of commitment: after integrating forward, can the supplier really commit to shuting off supply to rival downstream firms? On the second question, the classic argument against government intervention is the "one monopoly rent" story associated with the Chicago School. The basic argument is that, in some circumstances, a monopolist supplier can achieve the same profit regardless of whether it integrates forward and raises the downstream price or simply raises the input price. Whinston (2007) discusses a range of conditions under which this argument does not hold, and vertical integration or exclusive contracts indeed can be anticompetitive.

32. Of course, the same argument applies to any tactic that might generate market power, and the Teece (1986) proposal has become a general management doctrine for creating rents at the firm level. See, for example, the widely cited article by Amit and Schoemaker (1993).

alternatively formulate a nonmarket forward contract to assure supply.[33] So Carlton's theory provides a potential explanation in which vertical integration might be linked to the existence of a well-functioning market interface between layers of production (see also Green 1986).

A related observation can be made in the context of transaction cost analyses of integration. If we consider, for instance, the studies of transportation industries, such as trucking or shipping discussed above, a central problem being solved is logistical—the efficient short-run allocation of heterogeneous suppliers to heterogeneous tasks. The existence of market institutions—standard contractual forms, preexisting agreements on exactly what service will be exchanged, or the availability of comparable reference prices on which to base negotiation—can be crucial to support a disintegrated structure, because they are necessary to support efficient, predictable, and potentially competitive agreements at low cost.

The existence of market institutions to support disintegrated exchange is not a given. In many cases, market institutions tend to co-evolve with industry structure. Financial markets provide an exceptionally rich set of examples. Consider, for instance, a loan from a bank to an individual to allow the individual to purchase a house. Historically in the mortgage market, the same bank extended the loan and then undertook collections and received interest and principal payments from the borrower. In this setting, selling a loan portfolio would have been an idiosyncratic and unusual transaction. Yet over time, the development of relatively standard securitization contracts allowed the initial originators of loans (banks and later also less well-capitalized companies) to sell their loans to investors, creating a vertically disintegrated market structure, albeit one with some potential incentive problems on the part of loan originators.

The securitization example is useful: one can in fact point to a range of institutions required to support a disintegrated market structure. These include not only standardized contracts to facilitate sales of loans but also effective institutions for transferring foreclosure rights in the event of a default, rating agencies to limit the potential for adverse selection in the sale of loans, and functioning markets to allow investors who have bought securities to later sell them. From this perspective, standard transaction cost variables, such as complexity and specificity, look somewhat different. A particular transaction—say, the sale of a loan portfolio—might seem highly complex and specific at one point in time, but much less so once market institutions are developed to support it.[34]

The more general point here is that transaction costs depend on existing market institutions—institutions that facilitate search and matching, and institutions that facilitate contractual and pricing arrangements. As a result, a disintegrated market structure—particularly the creation of a disintegrated industry with frequent arm's-length exchanges—often requires the creation of market institutions: standards for products and contracts, and mechanisms for matching buyers and sellers, determining prices, disseminating information, and so forth. Although these issues fall outside the traditional theory of the firm, they become hard to avoid

33. Bolton and Whinston (1993) provide an alternative model, based on the property rights theory, that also shows how supply assurance motives can lead firms to vertically integrate.

34. At the same time, the creation of market institutions is facilitated by having a degree of standardization in the underlying products. It may be relatively easy to develop an effective market for loan resale if the loans themselves are relatively standardized in most dimensions.

when one takes the theory to empirical settings, and perhaps they deserve more attention from organizational economists interested in firm boundaries.

4.4. Coordination and Innovation

A wide range of economically important innovations depend on complementary but distinct inventions. Computers would not be particularly valuable without applications software, to take one famous example. New products are more valuable if process innovations allow them to be produced cheaply. The invention of mass production itself would have been much less valuable without transportation technologies that permitted large plants to serve geographically dispersed customers. Research on platform industries, on product and process innovations, and on general-purpose technologies all emphasize the importance of coordinating complementary invention and raise the question of whether this should take place within large organizations or across disparate innovative firms.[35] The resulting trade-offs provide an interesting example of how organizational theories of integration intersect with considerations of market structure.

A potential benefit of integrating the invention of complementary technologies is that investments may be better coordinated. A basic price theory intuition is that creators of complementary inventions may end up in a position of bilateral or multilateral monopoly, and they may fail to coordinate their pricing and other ex post decisions. This can lead to a form of hold-up where the full returns on ex ante investment are not appropriated. Kenneth Arrow (1974) famously pointed out one particular problem with contracting over innovation, which is that firms seeking to reach an agreement may find it difficult to exchange information in a way that protects their ideas. For these reasons, vertical integration, or a similar contract to internalize externalities in complementary invention, can increase the private return from innovative activity.

A second and distinct argument for integration comes from game theory. The multiple invention problem is a coordination game, and it might have failures to coordinate if the different inventors cannot be sure ex ante of one another's plans for invention. In addition, separate firms may have an incentive to delay investments in order to assess which technologies or standards will emerge as grounds for coordination. These forces work against decentralized innovation and favor ex ante partner selection and contracting.

A countervailing force comes from the economics of invention with uncertain goals or methods. Suppose that, in one or more of the complementary inventions, the highest value direction of technical progress is ex ante uncertain. Different potential inventors can have different views about the appropriate direction, based on their private information. In such circumstances, ex ante selection among alternative invention projects can be inefficient, even when there is a fixed cost of inventive effort.[36] The alternative is to have invention competition and then let the market choose the best alternative. The advantage of ex post invention market

35. In the case of truly infrastructural inventions, such as those in transportation, communications, or basic scientific research, government has also played a considerable role in coordinating innovation. In this chapter, however, we focus on the trade-off between markets and integrated firms.

36. See Nelson (1961), Evenson and Kislev (1976), Metcalfe (1986), and Klepper (1996) for models of competition with uncertainty about the direction of technical progress with ex post market selection.

selection is that it can choose the best innovative efforts out of a number of initiatives based on their technical success or their success in meeting market needs. As a result, a market form of organization may have considerable value when there is substantial uncertainty about the optimal direction of technical progress or about the source of that progress,

The computer industry, which over time has witnessed both a high degree of vertical integration and considerable disintegration, provides a useful case study of these considerations.[37] The first highly valuable segment of the industry, mainframe computers for corporate data processing, was pioneered in the 1950s. IBM, which dominated the market for four decades, was highly vertically integrated in invention and production. Its inventions included not only the computer but also a large number of complementary hardware and software components useful in business data processing.[38] As a result, the boundary of the firm coincided with the boundary between general and specific technologies. IBM sought to supply all general-purpose components for its corporate customers, including the tools that would let them invent specific applications.

What factors might have led to all of the general components, each of which needed ongoing invention, being supplied by one firm? The key was the ex ante identification of a desired direction of technical progress for the system as a whole. At the time of IBM's entry into computing, the basic technical knowledge associated with the computer itself was largely public.[39] IBM was already the leading seller of business data-processing products and subsequently built a research and invention capability in computers.[40] As a result, IBM had knowledge not only of the technical opportunity but also of demand needs. This information gave a single firm considerable ex ante knowledge about the set of feasible directions for improvements in the overall system and the set of products that would be valuable to its customers. IBM picked reasonable innovative goals from the intersection of these two sets, allowing it to coordinate multiple inventions in the general mainframe components—computer, tape drive, programming languages and operating system—going forward.[41]

37. See Bresnahan and Greenstein (1999) and Bresnahan and Malerba (2002). See also Langlois and Robinson (1992) on the value of vertical disintegration at the PC industry segment's founding.

38. These components included data-processing complements (e.g., tape or disk storage devices), input and output devices (e.g., printers), and especially software (e.g., operating systems and database management systems). Throughout IBM's tenure as the leading firm in this segment, customers had a high opinion of the way the complements making up an IBM system worked together but not of the technical level of its computers themselves.

39. It is important to point out that this discussion refers to the creation of computer systems for business use, not to the original invention of the computer itself. Many essential elements of the computer itself were invented before IBM's entry. The view, taken by IBM and others, that these essential elements were in the public domain, was extremely controversial. The controversy was heightened by IBM's late entry into the computer business and appropriation of existing knowledge. See Bresnahan and Malerba (2002) for a discussion. The point here depends on the fact that IBM, already knowledgeable about demand needs, could easily appropriate knowledge associated with the invention of the computer itself, not on the propriety of that appropriation.

40. The earlier business products were mechanical and electromechanical products, such as card sorters.

41. Industry historians will point out that this story misses important nuances. For example, the later emergence of applications software companies, especially semi-custom software developers such as EDS, meant that the distinction between the general and the specific is not such a bright line. Our point that the IBM system was largely one of vertical integration of the general components is robust, however.

The same trade-off worked the other way at the founding of another important computer industry segment, the personal computer. In that instance, ex ante knowledge about the overlap of technically feasible inventions and demand needs was much weaker, making coordination difficult. None of the founders of the personal computer industry—the designers of computers, disk drives, programming tools or operating systems—had a clear view of the appropriate direction of technical change ex ante. As a result, the original industry structure was not vertically disintegrated, with considerable independent innovation in each component. Coordination was achieved ex post invention through market institutions, such as voluntary compliance with interconnection standards.[42]

Eventually, the invention of certain key applications, such as the word processor and the spreadsheet, helped clarify the direction of technical change—namely, toward business applications. That increased clarity led to efforts to coordinate supply by contract, notably in the IBM PC.[43] Those contracts, however, did not specify the direction of technical change, nor were they exclusive on either side. Innovation in the different components continued to be organized more along market lines than by contract.[44] Indeed, some of the leading firms involved in IBM PC contracts, including IBM itself, were later replaced by others through market selection. The advantage of vertical disintegration of invention, even of the general-purpose components, was the possibility of component-by-component invention races and the extensive use of ex post market selection, also on a component-by-component basis.

This vertical industry structure and component-by-component market selection process continued through a number of rounds of fundamental innovation. Indeed, some dominant firms supplying important PC markets were replaced through competition for the market on a component-by-component basis.[45] As long as the overlap between technical opportunity and demand needs was hard to see ex ante the next round of innovation, the less-coordinated model had considerable advantage.

Another example of the trade-offs in coordinating innovative activity comes from the organization of scientific research. Here the relevant boundary is between firms that undertake research with the express purpose of commercialization, and universities and other organizations that undertake research with potentially broader objectives. The resulting industry structure often locates basic research in noncommercial organizations, with applied research taking place in firms in a vertically disintegrated structure (though the distinction is frequently blurred). In a pioneering article, Dasgupta and David (1997) emphasize two points about this structure: first, that the reward structure for academic scientists is very different than it is for commercial inventors; second, that noncommercial research institutions are organized very differently than are firms.

42. Once again, we skip important nuances. Some firms, most successfully Apple, had a more vertically integrated model. See Langlois and Robertson (1992) for more details, and for the observation that the more vertically integrated efforts failed precisely because of their integration.

43. IBM entered with a computer and sought to contract with the leading firm in each important complementary technology, successfully so in disk drives, spreadsheets, programmer tools, and microprocessors. In operating systems or word processors, the leading firm would not sign and IBM found other partners.

44. Even hardware specifications of the computer itself were set in the market by firms other than IBM. See Bresnahan and Greenstein (1999) for details.

45. See Langlois and Robertson (1992) and Bresnahan and Greenstein (1999) for details.

The structure of incentives facing scientists rewards them less with money and more with freedom, prestige, and status (Stern 2004). These peculiar goals are reflected in a large number of distinctive scientific institutions, including peer review, status based on priority rather than ultimate commercial success, and openness in publication and the permission of replication. Because these institutions were designed by scientists rather than by managers (i.e., individuals interested primarily in the commercialization of science), they lead to a situation in which university scientists' rewards can depend more on other scientists than on the managers in their own organizations. The incentives created by scientific institutions can enhance productivity—openness, for example, lowers the costs of access to existing knowledge by a large group of potential inventors, while competition to be first pushes each discipline toward more rapid technical progress. But it also can have disadvantages, for example, if peer review leads to rewards for work that is highly regarded by one's peers but not particularly useful or interesting to others outside one's narrow field. In contrast, the system of incentives for commercial innovation is more straightforward, as the market provides a primary metric for success.

These differences lead to tensions when basic and applied research take place in a disintegrated fashion. Commercially oriented managers experience difficulty incentivizing scientists to engage in work that is not rewarded by their discipline (e.g., work on specific rather than general problems) or to interact with others outside their field. For scientists who are considering commercializing their inventions or discoveries, there can be a commercial incentive to keep innovations secret rather than to publish, weakening the reward system of academic science (Murray and Stern 2007). And successful structures at the boundary of science and commercialization, such as university licensing programs kicked off by the Bayh-Dole Act permitting university scientists to patent inventions that result from government-funded research, sometimes have been criticized for distorting basic scientific research incentives.

Nevertheless, there are considerable reasons to think that a system whereby basic scientific research was left to commercial firms would be problematic. Perhaps the most obvious are the lack of immediate commercial reward to foundational research and the spillovers that result from many scientific breakthroughs. Even when scientific research produces inventions that are in principle appropriable, they may not be appropriable in their initially planned use. A large and diverse firm might be able to capture some spillovers from scientific research but not all of them. A famous example comes from Bell Laboratories. Bell Labs worked not only on basic engineering projects but also on basic science, discovering (among other things) the transistor effect. Although the Bell System certainly had uses for a wide range of inventions, the breadth of use for the transistor was too wide even for highly diverse Bell, so the firm put that fundamental innovation into the public domain as part of a strategy to avoid government interference.

5. Conclusion

We have discussed a wide range of economic theory and empirical work related to the vertical integration or disintegration of production. Theories from organizational economics have tended to focus on the contractual issues that affect the integration decisions of individual firms, posing an interesting contrast to resource- or capabilities-based theories of integration, and to industrial organization theories that emphasize the relationship between integration decisions and the determination of horizontal market structure. We have argued that the intersection of

these areas—in particular, how transaction cost issues interact with other forces to determine the overall structure of industries and markets—deserves more attention. We also see considerable room for progress in formulating empirical tests of the various theories and in measuring the incidence and importance of transaction costs under alternative contractual or organizational arrangements. As better data on contracts and organizations become available, most notably in emerging areas (e.g., the online world and rapidly industrializing economies), we expect there will be many opportunities along these lines.

REFERENCES

Acemoglu, Daron, Simon Johnson, and Todd Mitton. 2009. "Determinants of Vertical Integration: Financial Development and Contracting Costs." *Journal of Finance* 63: 1251–1290.

Acemoglu, Daron, Philippe Aghion, Rachel Griffith, and Fabrizio Zilibotti. 2010. "Vertical Integration and Technology: Theory and Evidence." *Journal of the European Economic Association* 8: 989–1033.

Afendulis, Christopher, and Daniel Kessler. 2007. "Tradeoffs from Integrating Diagnosis and Treatment in Markets for Health Care." *American Economic Review* 97: 1013–1020.

———. 2011. "Vertical Integration and Optimal Reimbursement Policy." Working Paper 17316, National Bureau of Economic Research, Cambridge, MA.

Alchian, Armen A., and Harold Demsetz. 1972. "Production, Information Costs, and Economic Organization." *American Economic Review* 62: 777–795.

Amit, Raphael, and Paul Schoemaker. 1993. "Strategic Assets and Organizational Rent." *Strategic Management Journal* 14: 33–46.

Anderson, Erin, and David C. Schmittlein. 1984. "Integration of the Sales Force: An Empirical Examination." *RAND Journal of Economics* 15: 385–395.

Anton, James J., and Dennis A. Yao. 1995. "Start-Ups, Spin-Offs and Internal Projects." *Journal of Law, Economics, and Organization* 11: 362–378.

Antras, Pol, and Esteban Rossi-Hansberg. 2009. "Organizations and Trade." *Annual Review of Economics* 1: 43–64.

Arora, Ashish, Andrea Fosfuri, and Alfonso Gambardella. 2001. *Markets for Technology: The Economics of Innovation and Corporate Strategy*. Cambridge, MA: MIT Press.

Arora, Ashish, Alfonso Gambardella, L. Magazzini, and Fabio Pammolli. 2009. "A Breath of Fresh Air? Firm Types, Scale, Scope and Selection Effects in Drug Development." *Management Science* 55: 1638–1653.

Arrow, Kenneth J. 1974. *The Limits of Organization*. New York: W. W. Norton.

———. 1975. "Vertical Integration and Communication." *Bell Journal of Economics* 6: 173–183.

Asanuma, Banri. 1989. "Manufacturer-Supplier Relationships in Japan and the Concept of Relation-Specific Skill." *Journal of the Japanese and International Economies* 3: 1–30.

Atalay, Enghin, Ali Hortaçsu, and Chad Syverson. 2012. "Why Do Firms Own Production Chains?" Working paper, University of Chicago.

Azoulay, Pierre. 2004. "Capturing Knowledge within and across Firm Boundaries: Evidence from Clinical Development." *American Economic Review* 94: 1591–1612.

Baker, George P., and Thomas N. Hubbard. 2003. "Make versus Buy in Trucking: Asset Ownership, Job Design and Information." *American Economic Review* 93: 551–572.

———. 2004. "Contractibility and Asset Ownership: On-Board Computers and Governance in U.S. Trucking." *Quarterly Journal of Economics* 119: 1443–1479.

Baker, George P., Robert Gibbons, and Kevin Murphy. 2011. "Relational Adaptation." Working paper, Massachusetts Institute of Technology, Cambridge, MA.

Baldwin, Carliss, and Kim Clark. 2005. *Design Rules,* volume 1: *The Power of Modularity*. Cambridge, MA: MIT Press.

Berle, Adolf, and Gardiner Means. 1932. *The Modern Corporation and Private Property*. New York: Macmillan.

Berry, Steven, and Peter Reiss. 2007. "Empirical Models of Entry and Market Structure," in Mark Armstrong and Robert Porter (eds.), *Handbook of Industrial Organization,* volume 3. Amsterdam: North-Holland, pp. 1845–1886.

Bloom, Nicholas, and John van Reenen. 2007. "Measuring and Explaining Management Practices across Firms and Countries." *Quarterly Journal of Economics* 122: 1351–1408.

Bolton, Patrick, and Michael Whinston. 1993. "Incomplete Contracts, Vertical Integration, and Supply Assurance." *Review of Economic Studies* 60: 121–148.

Bresnahan, Timothy, and Alfonso Gambardella. 1998. "The Division of Inventive Labor and the Extent of the Market," in Elhanan Helpman (ed.), *General Purpose Technologies and Economic Growth.* Cambridge, MA: MIT Press, pp. 253–281.

Bresnahan, Timothy, and Shane Greenstein. 1999. "Technological Competition and the Structure of the Computer Industry." *Journal of Industrial Economics* 47: 1–40.

Bresnahan, Timothy, and Franco Malerba. 2002. "The Value of Competitive Innovation," in Chong-En Bai and Chi-Wa Yuen (eds.), *Technology and the New Economy*. Cambridge, MA: MIT Press, pp. 49–94.

Bresnahan, Timothy, Rebecca Henderson, and Shane Greenstein. 2012. "Schumpeterian Competition and Diseconomies of Scope: Illustrations from the Histories of Microsoft and IBM," in Josh Lerner and Scott Stern (eds.), *The Rate and Direction of Technical Change Revisited.* Chicago: University of Chicago Press, in press.

AU: Any update yet? Please supply the page range for the chapter if available now.

Brickley, James A., and Frederick H. Dark. 1987. "The Choice of Organizational Form: The Case of Franchising." *Journal of Financial Economics* 18: 401–420.

Calvo, Guillermo, and Stanislaw Wellisz. 1978. "Supervision, Loss of Control and the Optimal Size of the Firm." *Journal of Political Economy* 86: 943–952.

Carlton, Dennis. 1979. "Vertical Integration in Competitive Markets under Uncertainty." *Journal of Industrial Economics* 27: 189–209.

Casadesus-Masanell, Ramon, and Daniel F. Spulber. 2000. "The Fable of Fisher Body." *Journal of Law and Economics* 43: 67–104.

Caves, Richard E., and William F. Murphy, II. 1976. "Franchising: Firms, Markets, and Intangible Assets." *Southern Economic Journal* 42: 572–586.

Ciliberto, Federico. 2006. "Does Organizational Form Affect Investment Decisions?" *Journal of Industrial Economics* 54: 63–93.

Ciliberto, Federico, and David Dranove. 2006. "The Effect of Physician-Hospital Affiliations on Hospital Prices in California." *Journal of Health Economics* 25: 29–38.

Clark, Kim. 1985. "The Interaction of Design Hierarchies and Market Concepts in Technological Evolution." *Research Policy* 14: 235–251.

Coase, Ronald H. 1937. "The Nature of the Firm." *Economica* 4: 386–405.

———. 1972. "Industrial Organization: A Proposal for Research," in Victor R. Fuchs (ed.), *Policy Issues and Research Opportunities in Industrial Organization*. New York: National Bureau of Economics Research, pp. 59–73.

———. 1988. "The Nature of the Firm: Influence." *Journal of Law, Economics, and Organization* 4: 33–47.

———. 2000. "The Acquisition of Fisher Body by General Motors." *Journal of Law and Economics* 43: 15–31.

Cohen, Wesley, and Daniel Levinthal. 1990. "Absorptive Capacity: A New Perspective on Learning and Innovation." Administrative Science Quarterly 35: 128–152.

Crocker, Keith, and Kenneth J. Reynolds. 1993. "The Efficiency of Incomplete Contracts: An Empirical Analysis of Air Force Engine Procurement." *RAND Journal of Economics* 24: 126–146.

Cuellar, Alison Evans, and Paul J. Gertler. 2006. "Strategic Integration of Hospitals and Physicians." *Journal of Health Economics* 25: 1–28.

Cyert, Richard, and James March. 1963. *A Behavioral Theory of the Firm.* Englewood Cliffs, NJ: Prentice-Hall.

Dasgupta, Partha, and Paul David. 1997. "Information Disclosure and the Economics of Science and Technology," in George Feiwel (ed.), *Arrow and the Ascent of Modern Economic Theory.* New York: New York University Press, pp. 539–591.

Einav, Liran, Amy Finkelstein, and Jonathan Levin. 2010. "Beyond Testing: Empirical Models of Insurance Markets." *Annual Review of Economics* 2: 311–336.

Elberfeld, Wallter. 2002. "Market Size and Vertical Integration: Stigler's Hypothesis Reconsidered." *Journal of Industrial Economics* 50: 23–43.

Evenson, Robert E., and Yoav Kislev. 1976. "A Stochastic Model of Applied Research." *Journal of Political Economy* 84: 265–281.

Forbes, Silke, and Mara Lederman. 2009. "Adaptation and Vertical Integration in the Airline Industry." *American Economic Review* 99: 1831–1849.

———. 2012a. "Does Vertical Integration Affect Firm Performance? Evidence from the Airline Industry." *RAND Journal of Economics* 41: 765–790.

AU: Any update yet? Please supply the volume number and page range if available.

———. 2012b. "Contract Form and Technology Adoption in a Network Industry." *Journal of Law, Economics, and Organization,* forthcoming.

Freeland, Robert F. 2000. "Creating Holdup through Vertical Integration: Fisher Body Revisited." *Journal of Law and Economics* 43: 33–66.

Garicano, Luis. 2000. "Hierarchies and the Organization of Knowledge in Production." *Journal of Political Economy* 108: 874–904.

Gaynor, Martin. 2006. "Is Vertical Integration Anticompetitive? Definitely Maybe (But That's Not Final)." *Journal of Health Economics* 25: 175–180.

Gibbons, Robert. 2005. "Four Formal(izable) Theories of the Firm?" *Journal of Economic Behavior and Organization* 58: 200–245.

Gil, Ricard. 2007. "'Make-or-Buy' in Movies: Integration and ex-post Renegotiation." *International Journal of Industrial Organization* 25: 643–655.

Goldberg, Victor. 1976. "Regulation and Administered Contracts." *Bell Journal of Economics* 7: 426–448.

Green, Jerry. 1986. "Vertical Integration and Assurance of Markets," in G. Frank Mathewson and Joseph E. Stiglitz (eds.), *New Developments in the Analysis of Market Structure.* Cambridge, MA: MIT Press, pp. 177–207.

Grossman, Sanford, and Oliver Hart. 1986. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy* 94: 691–719.

Hart, Oliver. 1995. *Firms, Contracts, and Financial Structure.* Clarendon Lectures in Economics. Oxford and New York: Oxford University Press and Clarendon Press.

———. 2009. "Hold-Up, Asset Ownership, and Reference Points." *Quarterly Journal of Economics* 124: 267–300.

Hart, Oliver, and Bengt Holmström. 2010. "A Theory of Firm Scope." *Quarterly Journal of Economics* 125: 483–513.

Hart, Oliver, and John Moore. 1990. "Property Rights and the Nature of the Firm." *Journal of Political Economy* 98: 1119–1158.

Hart, Oliver, and Jean Tirole. 1990. "Vertical Integration and Market Foreclosure." *Brookings Papers on Economic Activity: Microeconomics* 1990: 205–276.

Helpman, Elhanan. 2006. "Trade, FDI, and the Organization of Firms." *Journal of Economic Literature* 44: 589–630.

Henderson, Rebecca, and Iain Cockburn. 1996. "Scale, Scope, and Spillovers: The Determinants of Research Productivity in Drug Discovery." *RAND Journal of Economics* 27: 32–59.

Hennart, Jean-Francois. 1988. "A Transaction Costs Theory of Equity Joint Ventures." *Strategic Management Journal* 9: 361–374.

Holmes, Thomas. 1999. "Localization of Industry and Vertical Disintegration." *Review of Economics and Statistics* 81: 314–325.

Holmström, Bengt. 1999. "The Firm as a Subeconomy." *Journal of Law, Economics, and Organization* 15: 74–102.

Holmström, Bengt, and Paul Milgrom. 1994. "The Firm as an Incentive System." *American Economic Review* 84: 972–991.

Holmström, Bengt, and John Roberts. 1998. "The Boundaries of the Firm Revisited." *Journal of Economic Perspectives* 12: 73–94.

Hortaçsu, Ali, and Chad Syverson. 2007. "Cementing Relationships: Vertical Integration, Foreclosure, Productivity, and Prices." *Journal of Political Economy* 115: 250–301.

John, George, and Barton A. Weitz. 1988. "Forward Integration into Distribution: An Empirical Test of Transaction Cost Analysis." *Journal of Law, Economics, and Organization* 4: 337–355.

Joskow, Paul L. 1985. "Vertical Integration and Long-Term Contracts: The Case of Coal-Burning Electric Generating Plants." *Journal of Law, Economics, and Organization* 1: 33–80.

———. 1987. "Contract Duration and Relationship Specific Investments." *American Economic Review* 77: 168–175.

———. 1988. "Price Adjustment in Long Term Contracts: The Case of Coal." *Journal of Law and Economics* 31: 47–83.

———. 2005. "Vertical Integration," in Claude Ménard and Mary Shirley (eds.), *Handbook of New Institutional Economics*. Dordrecht and New York: Springer, pp. 319–348.

Kehoe, Michael R. 1996. "Franchising, Agency Problems, and the Cost of Capital." Applied Economics 28: 1485–1493.

Kerkvliet, Joe. 1991. "Efficiency and Vertical Integration: The Case of Mine-Mouth Electric Generating Plants." *Journal of Industrial Economics* 39: 467–482.

Klein, Benjamin. 1988. "Vertical Integration as Organized Ownership: The Fisher Body–General Motors Relationship Revisited." *Journal of Law, Economics, and Organization* 4: 199–213.

———. 1996. "Why Hold-Ups Occur: The Self-Enforcing Range of Contractual Relationships." *Economic Inquiry* 34: 444–463.

———. 2000. "Fisher–General Motors and the Nature of the Firm." *Journal of Law and Economics* 43: 105–141.

———. 2007. "The Economic Lessons of Fisher Body–General Motors." *International Journal of the Economics of Business* 14: 1–36.

Klein, Benjamin, Robert Crawford, and Armen Alchian. 1978. "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process." *Journal of Law and Economics* 21: 297–326.

Klein, Peter G. 2005. "The Make-or-Buy Decision: Lessons from Empirical Studies," in Claude Ménard and Mary Shirley (eds.), *Handbook of New Institutional Economics*. Dordrecht and New York: Springer, pp. 435–464.

Klepper, Steven. 1996. "Entry, Exit, Growth, and Innovation over the Product Life Cycle." *American Economic Review* 86: 562–583.

Lafontaine, Francine. 1992. "Agency Theory and Franchising: Some Empirical Results." *RAND Journal of Economics* 23: 263–283.

Lafontaine, Francine, and Scott Masten. 2002. "Contracting in the Absence of Specific Investments of Moral Hazard: Understanding Carrier-Driver Relations in the US." Working Paper 8859, National Bureau of Economic Research, Cambridge, MA.

Lafontaine, Francine, and Kathryn L. Shaw. 2005. "Targeting Managerial Control: Evidence from Franchising." *RAND Journal of Economics* 36: 131–150.

Lafontaine, Francine, and Margaret E. Slade. 2007. "Vertical Integration and Firm Boundaries: The Evidence." *Journal of Economic Literature* 45: 629–685.

Langlois, Richard, and Paul L. Robertson. 1989. "Explaining Vertical Integration: Lessons from the American Automobile Industry." *Journal of Economic History* 69: 361–375.

———. 1992. "Networks and Innovation in a Modular System: Lessons from the Microcomputer and Stereo Component Industries." *Research Policy* 21: 297–313.

Levin, Jonathan, and Steve Tadelis. 2010. "Contracting for Government Services: Theory and Evidence from U.S. Cities." *Journal of Industrial Economics* 58: 507–541.

Levy, D. T. 1984. "Testing Stigler's Interpretation of 'Division of Labor Is Limited by the Extent of the Market.'" *Journal of Industrial Economics* 32: 377–389.

Lieberman, Marvin B. 1991. "Determinants of Vertical Integration: An Empirical Test." *Journal of Industrial Economics* 39: 451–466.

Lucas, Robert. 1978. "On the Size Distribution of Business Firms." *Bell Journal of Economics* 9: 508–523.

Malmgren, Harald B. 1961. "Information, Expectations and the Theory of the Firm." *Quarterly Journal of Economics* 75: 399–421.

Masten, Scott E. 1984. "The Organization of Production: Evidence from the Aerospace Industry." *Journal of Law and Economics* 27: 403–417.

Masten, Scott E., James W. Meehan, Jr., and Edward A. Snyder. 1989. "Vertical Integration in the U.S. Auto Industry: A Note on the Influence of Transaction Specific Assets." *Journal of Economic Behavior and Organization* 12: 265–273.

———. 1991. "The Costs of Organization." *Journal of Law, Economics, and Organization* 7: 1–25.

McAfee, Preston, and John McMillan. 1995. "Organizational Diseconomies of Scale." *Journal of Economics and Management Strategy* 4: 399–426.

Metcalfe, J. S. 1986. "Technological Variety and the Process of Competition." *Economie appliquée* 39: 493.

Minkler, Alanson P. 1990. "An Empirical Analysis of a Firm's Decision to Franchise." *Economics Letters* 34: 77–82.

Monteverde, Kirk, and David J. Teece. 1982. "Supplier Switching Costs and Vertical Integration in the Automobile Industry." *Bell Journal of Economics* 13: 206–213.

Mullainathan, Sendhil, and David Scharfstein. 2001. "Do Firm Boundaries Matter?" *American Economic Review* 91: 195–199.

Murray, Fiona, and Scott Stern. 2007. "When Ideas Are Not Free: The Impact of Patenting on Academic Science," in Adam Jaffe, Josh Lerner, and Scott Stern (eds.), *Innovation Policy and the Economy,* volume 7. Cambridge, MA: National Bureau of Economic Research, pp. 33–69.

Nelson, Richard R. 1961. "Development Efforts." Review of Economics and Statistics 43: 351–364.

Nickerson, Jack A., and Brian S. Silverman. 2003. "Why Aren't All Truck Drivers Owner-Operators? Asset Ownership and the Employment Relation in Interstate For-Hire Trucking." *Journal of Economics and Management Strategy* 12: 91–118.

Norton, Seth W. 1988. "An Empirical Look at Franchising as an Organizational Form." *Journal of Business* 61: 197–218.

Novak, Sharon, and Scott Stern. 2009. "Complementarity among Vertical Integration Decisions: Evidence from Automobile Product Development." *Management Science* 54: 1963–1979.

Ohanian, Nancy. 1994. "Vertical Integration in the US Pulp and Paper Industry, 1900–1940." *Review of Economics and Statistics* 76: 202–207.

Penrose, Edith. 1959. *The Theory of the Growth of the Firm*. New York: John Wiley and Sons.

Perry, Martin K. 1989. "Vertical Integration," in Richard Schmalensee and Robert Willig (eds.), *Handbook of Industrial Organization*. Amsterdam: North-Holland, pp. 183–258.

Pirrong, Stephen Craig. 1993. "Contracting Practices in Bulk Shipping Markets: A Transactions Cost Explanation." *Journal of Law and Economics* 36: 937–976.

Radner, Roy. 1992. "Hierarchy: The Economics of Managing." *Journal of Economic Literature* 30: 1382–1415.

Ravenscraft, David, and Frederic Scherer. 1987. *Mergers, Sell-Offs and Economic Efficiency*. Washington, DC: Brookings Institution.

Regan, Laureen. 1999. "Vertical Integration in the Property-Liability Insurance Industry: A Transaction Cost Approach." *Journal of Risk and Insurance* 64: 41–62.

Rosenberg, Nathan. 1994. *Exploring the Black Box: Technology, Economics and History*. Cambridge: Cambridge University Press.

———. 1998. "Chemical Engineering as a General Purpose Technology," in Elhanan Helpman (ed.), *General Purpose Technologies and Economic Growth*. Cambridge, MA: MIT Press, pp. 167–192.

Schoar, Antoinette. 2002. "Effects of Corporate Diversification on Productivity." *Journal of Finance* 57: 2379–2403.

Shelanski, Howard A., and Peter G. Klein. 1995. "Empirical Research in Transaction Cost Economics: A Review and Assessment." *Journal of Law, Economics, and Organization* 11: 335–361.

Shepard, Andrea. 1993. "Contractual Form, Retail Price, and Asset Characteristics in Gasoline Retailing." *RAND Journal of Economics* 24: 58–77.

Simon, Herbert A. 1947. *Administrative Behavior*. New York: Macmillan.

———. 1951. "A Formal Theory of the Employment Relationship." *Econometrica* 19: 293–305.

Slade, Margaret E. 1996. "Multitask Agency and Contract Choice: An Empirical Exploration." *International Economic Review* 37: 465–486.

Spengler, Joseph J. 1950. "Vertical Integration and Antitrust Policy." *Journal of Political Economy* 58: 347–352.

Stein, Jeremy. 1997. "Internal Capital Markets and the Competition for Corporate Resources." *Journal of Finance* 52: 111–133.

Stern, Scott. 2004. "Do Scientists Pay to Be Scientists?" *Management Science* 50: 835–853.

Stigler, George. 1951. "The Division of Labor is Limited by the Extent of the Market." *Journal of Political Economy* 59: 185–193.

Stuckey, John. 1983. *Vertical Integration and Joint Ventures in the Aluminum Industry*. Cambridge, MA: Harvard University Press.

Sutton, John. 2007. "Market Structure: Theory and Evidence," in Mark Armstrong and Robert Porter (eds.), *Handbook of Industrial Organization,* volume 3. Amsterdam: North-Holland, pp. 2301–2368.

Teece, David. 1986. "Profiting from Technological Innovation: Implications for Integration, Collaboration, Licensing and Public Policy." *Research Policy* 15: 285–305.

Tirole, Jean. 1988. *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.

Van Biesebroeck, Johannes. 2006. "Complementarities in Automobile Production." Working Paper 12131, National Bureau of Economic Research, Cambridge, MA.

Walker, Gordon, and David Weber. 1984. "A Transaction Cost Approach to Make-or-Buy Decisions." *Administrative Science Quarterly* 29: 373–391.

Wernerfelt, Berger. 1984. "A Resource-Based View of the Firm." *Strategic Management Journal* 5: 171–180.
Whinston, Michael. 2003. "On the Transaction Cost Determinants of Vertical Integration." *Journal of Law, Economics, and Organization* 19: 1–23.
———. 2007. *Lectures on Antitrust Economics.* Cambridge, MA: MIT Press.
Williamson, Oliver E. 1971. "The Vertical Integration of Production: Market Failure Considerations." *American Economic Review* 61: 112–123.
———. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications.* New York: Free Press.
———. 1979. "Transaction-Cost Economics: The Governance of Contractual Relations." *Journal of Law and Economics* 22: 233–261.
———. 1985. *The Economic Institutions of Capitalism.* New York: Free Press.
———. 2005. "The Economics of Governance." *American Economic Review* 95: 1–18.
———. 2010. "Transaction Cost Economics: The Natural Progression." *American Economic Review* 100: 673–690.
Yeap, Clarissa A. 2006. "Residual Claims and Incentives in Restaurant Chains." Center for Economic Studies Working Paper CES 06–18, U.S. Census Bureau, Washington, DC.
Young, Allyn. 1928. "Increasing Returns and Economic Progress." *Economic Journal* 38: 527–542.